现代应用数学手册

《现代应用数学手册》编委会

概率统计与随机过程卷



清华大学出版社
http://www.tup.tsinghua.edu.cn

《现代应用数学手册》编委会

现代应用数学手册

概率统计与随机过程卷



清 华 大 学 出 版 社

(京)新登字 158 号

内 容 简 介

本书为介绍概率统计与随机过程之数学理论及应用方法的工具书. 全书分三大部分共 35 章. 第一部分为第 1 至 10 章, 介绍概率论中的基本概念和基本结论; 第二部分为第 11 至 20 章, 介绍各种经典和现代的数理统计方法; 第三部分为第 21 至 35 章, 介绍随机过程的一般理论和应用概率方法, 以及位势理论、鞅论、随机微分方程、预报与滤波、排队论、可靠性理论、随机模拟与马尔可夫决策规划等较新内容. 全书覆盖面广, 内容新颖, 实用性强, 查阅方便. 适合于广大科技人员、管理干部与大中专院校的师生.

图书在版编目(CIP)数据

现代应用数学手册. 概率统计与随机过程卷/《现代应用数学手册》编委会编. --北京: 清华大学出版社, 1999

ISBN 7-302-03553-9

Ⅰ. 现… Ⅱ. 现… Ⅲ. ①应用数学-手册②概率论③数理统计④随机过程 Ⅳ. 029-62

中国版本图书馆 CIP 数据核字(1999)第 14322 号

出版者: 清华大学出版社(北京清华大学学研楼, 邮编 100084)

http://www.tup.tsinghua.edu.cn

印刷者: 清华大学印刷厂

发行者: 新华书店总店北京发行所

开 本: 850×1168 1/32 **印张**: 28.5 **字数**: 735 千字

版 次: 2000 年 7 月第 1 版 2000 年 7 月第 1 次印刷

书 号: ISBN 7-302-03553-9/O・213

印 数: 0001～4000

定 价: 48.00 元

《现代应用数学手册》编辑委员会

概率统计与随机过程卷

责任编委　陆　璇　葛余博

章　次	编　者	校　者
1～6	汪国柄	陆　璇
7～10	陆　璇	葛余博
11～14	陆　璇	余　桂
15～17	余　桂	陆　璇
18	陆　璇	余　桂
19	周方俊	陆　璇
20	余　桂	陆　璇
21	董晓澎　葛余博	葛余博
22	杜神甫　葛余博	葛余博
23	李仲莲　赵衡秀	葛余博
24	陈典发	吴　荣
25～26	葛余博	刘秀芳
27	陈典发　李援南	吴　荣　葛余博
28	陈典发	吴　荣
29～30	葛余博	严加安
31	李仲莲	葛余博
32	陈宝谦	徐光辉
33	程　侃	曹晋华
34	林元烈	侯炳辉
35	林元烈	张道智

序

随着计算机科学技术的飞速发展，人类正进入信息时代.

信息时代是应用数学大发展的时代，人类长期积累起来的知识体系，正面临着第3次数学化. 数学思想，数学方法与数学模型随着计算机的广泛应用，日益渗透到各种行业中去.

当代，除了古典的数学理论（初等数学，微积分学，微分方程，复变函数等）早已得到广泛的应用外，一些比较抽象的现代数学理论（集合论、数理逻辑、范畴论、抽象代数、泛代数、代数几何、拓扑学、泛函分析等）以及一些新兴的数学理论（随机过程、时间序列、运筹学、最优化理论、有限元方法、模糊数学、混沌与分形等）也逐渐地成为社会生产，科学实验，工程技术及经济管理中不可缺少的工具，应用数学的适用范围正在迅速地扩大.

为了满足日益增长的社会需求，清华大学应用数学系《现代应用数学手册》编委会，组织编写了这套多卷集的手册.

本书读者是理、工、医、农、经管等各个领域中的广大工程技术人员、科研人员，大、中专院校的教师、学生、研究生及其他使用数学工具的实际工作者. 其中有些内容对于中学生也是适用的.

编者力求使本书成为一套高质量的工具书，它有下列特点：

(1) **内容"新颖"**　本书力求做到内容现代化，除用现代观点介绍古典内容外，对已出现的新理论、新方法尽量优先选入.

(2) **突出"应用"**　本书在选材上突出数学理论的应用，以通

俗易懂的方式着重介绍在现代科学技术等实际领域中应用广泛的数学理论和方法.

(3) **紧密"结合"计算机应用** 为了更有效地应用数学方法解决各种实际问题，广大科技人员迫切要求数学方法与计算机应用相结合，提高工作效率. 为此，本书在结合计算机应用方面，给予特别的重视.

(4) **版面设计"合理"，便于迅速查阅** 为方便读者使用，本书采用了一套较为完善的索引体系. 除正文中章、节的编号沿用国际通行的十进制编号外，对于重要的定义、定理、例题、公式、图、表等均有编号. 读者可以从 (1) 目录，(2) 中文—外文索引，(3) 外文—中文索引等三种途径，迅速找到所需资料. 此外，本书对载入的外国科学家人名，尽量采用"名从主人"的原则.

(5) **数学符号力求"统一"与国际化** 鉴于目前国内各种文献、书籍中使用的数学符号不够统一与国际化，增加了读者阅读时的困难. 本书除按国家标准 GB3102-93 外，兼用国际数学界权威著作《数学大百科辞典》(Encyclopedic Dictionary of Mathematics, EDM) 中的符号为标准. 对于不在上述文献中的其他新符号，则选用较为流行者.

本手册各卷内容独立完整，便于个人读者与团体读者按需选购. 当前应用数学急剧发展，编委会在条件成熟的时候，还将增出新卷.

本书的编撰是与清华大学应用数学系领导，特别是萧树铁教授的热心支持，编辑委员会各位编委的通力协作，校内外的许多教师、科研工作者的大力支持分不开的，编者深致谢意.

在编辑出版过程中，还得到清华大学出版社的热情支持.

本书从编撰到出版，历尽艰辛．饮水思源，编者还要感谢本书的发起人，清华大学应用数学系陆璇教授，北京出版社李利军编辑及已故的北京出版社社长王政人先生．

最后，编者还要对夫人王华敏表示谢忱，没有她的深刻理解、热情支持与持久的帮助，本书也难以问世．

主编　马振华

1997 年于清华园

编 者 说 明

近年来，在发达国家中，概率统计与随机过程的应用在科学研究、技术开发、生产管理与社会经济生活的各个领域内发挥了重要作用. 例如日本从50年代起将统计方法与工业质量管理结合起来. 对推动经济飞速发展起了至关重要的作用. 在我国，概率论与统计方法的应用也日益引起重视. 我们编写本书的目的就是为向广大科技工作者、管理人员与大中专院校师生提供一本概率与统计的工具书，以供查阅、使用.

本书包含三大部分. 第一部分内容为概率论基础，包括第1至10章，第二部分内容为数理统计方法，包括第11至20章，第三部分为随机过程与应用概率方法，包括第21至35章，本书的特点是覆盖面宽，内容先进，实用性强及便于查阅. 在编写过程中，我们力求对内容的叙述做到深入浅出，重点放在便于理解与应用上.

限于我们的水平，在编写中可能出现错误，欢迎广大读者给予指正.

编者

数学符号表

符号	含义
$\forall$	全称量词
$\exists$	存在量词
$\Leftrightarrow$	等价
$\triangleq$/$\hat{=}$	定义
$\max\{x,y\}/x \vee y$	最大值
$\min\{x,y\}/x \wedge y$	最小值
$\limsup/\overline{\lim}$	上极限
$\liminf/\underline{\lim}$	下极限
$\nabla f/\mathrm{grad} f$	f 的梯度
$\Delta f/\nabla^2 f$	拉普拉斯算子
$\mathbf{I}_n$	n 阶单位阵
$\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$	对角阵
$\lvert\boldsymbol{A}\rvert/\det\boldsymbol{A}$	方阵 $\boldsymbol{A}$ 的行列式
$\mathrm{tr}(\boldsymbol{A})$	方阵 $\boldsymbol{A}$ 的迹($\boldsymbol{A}$ 的对角元素之和)
$\mathrm{rk}(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的秩
$\mathbf{1}_n$	所有分量为 1 的 n 维向量
r.v.	随机变量的缩写
$F_{X\mid Y}(x\mid y)/F(x\mid y)$	给定 $Y=y$ 时 X 的条件分布函数
$f_{X\mid Y}(x\mid y)/f(x\mid y)$	给定 $Y=y$ 时 X 的条件密度函数
$E(X\mid Y)/E(X\mid Y=y)$	给定 $Y=y$ 时 X 的条件期望
$\sigma^2(X)/\mathrm{var}(X)/D(X)/V(X)/V$	随机变量 X 的方差

$\text{cov}(X,Y)$	随机变量 X 与 Y 的协方差
$\rho(X,Y)/\text{cor}(X,Y)$	随机变量 X 与 Y 的相关系数
$E(\boldsymbol{X})$	随机向量 $\boldsymbol{X}$ 的均值向量
$V(\boldsymbol{X})$	随机向量 $\boldsymbol{X}$ 的方差向量
$\text{cov}(\boldsymbol{X},\boldsymbol{Y})$	随机向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的协方差阵
$\text{cor}(\boldsymbol{X},\boldsymbol{Y})$	随机向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的相差阵
$B(n,p)$	参数为 n,p 的二项分布
$b(n,p;k)$	$B(n,p)$取 k 的概率值
$P(\lambda)$	参数为 λ 的泊松分布
$p(\lambda;k)$	$P(\lambda)$取 k 的概率值
$N(\mu,\sigma^2)$	均值为 μ，方差为 σ^2 的正态分布
$N(0,1)$	标准正态分布
$\Phi(x)$	标准正态分布的分布函数
$\varphi(x)$	标准正态分布的密度函数
$\langle\Omega,\mathscr{F}\rangle$	由空间 Ω 及 σ 域 $\mathscr{F}$ 构成的可测空间
$\langle\Omega,\mathscr{F},\mu\rangle$	由空间 Ω，σ 域 $\mathscr{F}$ 及测度 μ 构成的测度空间
$\langle\Omega,\mathscr{F},P\rangle/\text{p.s}$	概率空间
$\mathscr{B}$	博雷尔可测集类/博雷尔可测函数类
$\langle R,\mathscr{B}\rangle$	一维博雷尔空间
$\langle R^n,\mathscr{B}^n\rangle/\langle R_n,\mathscr{B}_n\rangle$	n 维博雷尔空间
$\langle R^\infty,\mathscr{B}^\infty\rangle$	无穷维博雷尔空间
$I_A(\cdot)$	A 的示性函数
$\int_\Omega X\text{d}\mu\Big/\int X\text{d}\mu$	X 在空间 Ω 上关于测度 μ 的积分
$\int_A X\text{d}\mu$	X 在可测集 A 上关于测度 μ 的积分
X^+	函数 X 的正部

X^-	函数 X 的负部
$\xrightarrow{\text{a.c.}}$ / $\xrightarrow{\text{a.s.}}$	几乎处处收敛
$\xrightarrow{\text{P}}$	依概率收敛
$\xrightarrow{\text{d}}$ / $\mathscr{L}$	依分布收敛
$\xrightarrow{\text{m.s.}}$	均方收敛
$\xrightarrow{L^p}$	按 L^p 空间距离收敛
l.i.m	均方收敛极限
$(X_1,\cdots,X_n)$	样本
$(x_1,\cdots,x_n)$	样本$(X_1,\cdots,X_n)$的观测值
$L(\theta)/L$	似然函数
$\beta(\theta)$	功效函数
$(X_{(1)},\cdots,X_{(n)})$	由样本$(X_1,\cdots,X_n)$生成的顺序统计量
$R_1,\cdots,R_n$	由样本$(X_1,\cdots,X_n)$生成的秩统计量
$L_n(m^k)$	k 个 m 水平的因子做 n 次试验的正交表
$L_n(m_1^{k_1}\times m_2^{k_2})$	k_1 个 m_1 水平的因子与 k_2 个 m_2 水平的因子做 n 次试验的正交表
$X_t\in\mathscr{F}_t$	X_t 为 $\mathscr{F}_t$ 可测函数
$L^2(\Omega)$	Ω 上平方可积函数空间
$L^p(\Omega)$	Ω 上 p 次可积函数空间
L 变换	拉普拉斯变换
LS 变换	拉普拉斯-斯蒂尔切斯变换
$\mathscr{F}_t$	σ 代数 $\mathscr{F}$ 的某个子 σ 代数
i.i.d.	独立同分布
h_A	首中 A 的时刻
∂A	集 A 的边界
$S(l,x)$	$\triangleq \lim\limits_{c\downarrow l} S[c,x]$

$\Sigma(l)$	$\triangleq \int_l^x M[y,x]\mathrm{d}s(y)$
$N(l)$	$\triangleq \int_l^x S[y,x]\mathrm{d}m(y) = \int_l^x M[l,y]\mathrm{d}s(y)$
	$\triangleq \lim\limits_{a \downarrow l} M[a,x]$
N_T	更新过程
$\{P_k^{*i}\}_{k\in N}$	$\{P_k\}_{k\in N}$的 i 重卷积(褶积)
$M \perp\!\!\!\perp N$	M,N 正交
$x^{\perp}$	x 的正交补
(M,N)	M,N 的内积
$\langle M,N\rangle$	M,N 的相互特征
$SD[E]$	(伊藤)随机微分[方程]
$SI[E]$	(伊藤)随机积分[方程]
$I(f,\omega)$	f 的伊藤积分
$M/M/1$	单个服务员排队系统
$M/M^{[X]}/1$	成批服务系统
FTA	故障树分析
π	策略
f^{∞}	平稳策略
$Z(\pi)=\{(Y_t,\Delta_t),t\in\mathbf{N}\}$	马尔可夫决策过程
$\{R_t(\tau)=r(Y_t,\Delta_t),t\in\mathbf{N}\}$	报酬过程

1 随机事件及其概率

随机事件及其概率是概率论中两个最基本的概念.本章从直观的角度来讨论这两个概念,在第7章讨论严格的数学形式.

1.1 随机事件

1.1.1 随机试验与随机事件

随机试验(random experiment)是一个可观察结果的人工或自然的过程,其产生的结果可能不止一个,且不能事先确定会产生什么结果.

随机试验是一个相当广泛的概念.例如,一次科学试验,一个受到观测记录的自然过程(如天气)或社会过程(如市场),一组数据的采集(如人口调查)等,都可以看作是一个随机试验.

样本空间(sample space)是一个随机试验的全部可能出现的结果的集合,通常记作 Ω. Ω 中的点(即一个可能出现的试验结果)称为**样本点**(sample point),通常记作 ω.

随机事件(random event)是一个随机试验的一些可能结果的集合,是样本空间的一个子集.它常用大写拉丁字母 $A,B,C,\cdots$ 表示.

以下简称随机事件为事件.在实际问题中,事件常用一句话来描述(见例1.1.1~1.1.4).当试验的结果(即一个样本点)属于某事件所对应的子集时,则称该事件发生.为了数学表示上的方便,空集也算作一个事件,记作 $\varnothing$. 样本空间 Ω 本身也表示一个事件.

例 1.1.1 将一枚硬币连掷两次,观察硬币落地后是花面向

上还是字面向上. 这是一个随机试验. 用 H 记花面向上, W 记字面向上, 则共有 4 个可能出现的结果(样本点):

$$\omega_1 = HH, \omega_2 = HW, \omega_3 = WH, \omega_4 = WW.$$

样本空间 $\Omega=\{\omega_1, \omega_2, \omega_3, \omega_4\}$.

其中 $\omega_1=HH$, 表示第一次与第二次均出现花面; $\omega_2=HW$, 表示第一次出现花面, 第二次出现字面; 等等.

考虑下述事件:

A="花面字面各出现一次",

B="第一次出现花面",

C="至少出现一次花面",

D="至多出现一次花面".

则

$A=\{\omega_2, \omega_3\}, B=\{\omega_1, \omega_2\}, C=\{\omega_1, \omega_2, \omega_3\}, D=\{\omega_2, \omega_3, \omega_4\}$.

例 1.1.2 记录某电话交换台在上午 9 点至 10 点之间内接到呼唤的次数, 是一次随机试验, 其试验结果可以用任一个非负整数值来表示. 因此样本空间 $\Omega=\{0,1,2,\cdots\}$. 设事件 A 为"接到 5 次以上的呼唤", 事件 B 为"没有接到呼唤". 则

$$A = \{5,6,\cdots\}, \quad B = \{0\}.$$

例 1.1.3 从一大批灯泡中随意抽一个检验其寿命(接通电源直到灯丝烧断为止), 这也是一个随机试验. 以小时为单位, 灯泡的寿命可以用任一个非负实数来表示, 样本空间 $\Omega=\{\omega \mid 0\leqslant\omega\}$. 设事件 A 为"灯泡寿命不超过 10000 小时", 则 $A=\{\omega \mid 0\leqslant\omega\leqslant 10000\}$.

例 1.1.4 为调查儿童健康状况, 从某地区某年龄组的儿童中随机地抽取 100 名称量其体重, 这也是一次随机试验. 以 kg 为单位, 则一个试验结果可以用 100 个非负实数来表示, 记为 $x_1, x_2, \cdots, x_{100}$. 样本空间为 $\Omega=\{\omega=(x_1,\cdots,x_{100}) \mid x_i\geqslant 0, i=1,\cdots,100\}$, 设事件 A 为"测试儿童平均体重超过 15kg", 则

$$A=\left\{\omega=(x_1,\cdots,x_{100})\,|\,x_i\geqslant 0,i=1,\cdots,100,\frac{\sum_{i=1}^{100}x_i}{100}\geqslant 15\right\}.$$

从以上几例可以看出,随机试验的结果是非常不同的,可以是一个数(例 1.1.2,1.1.3),也可以是一个数组(例 1.1.4)或是一个符号串(例 1.1.1).在实际问题中,要正确地认清样本空间和事件是如何构成的.

1.1.2 事件间的关系与运算

当谈到一个以上的事件时,总是对同一试验的样本空间上的事件而言的.

定义 1.1.5 两个事件 A 与 B 可能有以下几种特殊的关系:

(1) **包含**(contain) 若事件 B 发生则事件 A 也发生,称"A 包含 B"或"B 含于 A",记作 $A\supset B$ 或 $B\subset A$.

(2) **等价**(equivalent) 若 $A\supset B$ 且 $B\supset A$,即 A 与 B 同时发生或同时不发生,则称 A 与 B 等价,记作 $A=B$.

(3) **互斥**(exclusive) 若 A 与 B 不能同时发生,则称 A 与 B 互斥,记作 $AB=\varnothing$.

(4) **对立**(contrary) 若 A 与 B 互斥,且必有一个发生,则称 A 与 B 对立,记作 $A=\overline{B}$ 或 $B=\overline{A}$.又称 A 为 B 的**余事件**(complement)或 B 为 A 的**余事件**.

任意两个事件不一定会是上述几种关系中的一种.

由给定的一些事件按下述运算可导出新的事件.

定义 1.1.6 设 $A,B,A_1,A_2,\cdots,A_n$ 为一些事件,它们有下述的运算:

(1) **交**(intersection) 记 $C=$"A 与 B 同时发生",称为事件 A 与 B 的交,$C=\{\omega|\omega\in A$ 且 $\omega\in B\}$,记作 $C=A\cap B$,或 $C=AB$.

类似地用 $\bigcap_{i=1}^{n}A_i=A_1\cap A_2\cap\cdots\cap A_n=A_1A_2\cdots A_n$ 表示事件"几

个事件 $A_1,A_2,\cdots,A_n$ 同时发生".

$$\bigcap_{i=1}^{n} A_i = \{\omega|\omega \in A_i, i = 1,\cdots,n\}.$$

(2) **并**(union) 记 C="A 与 B 中至少有一个发生",称为事件 A 与 B 的并,$C=\{\omega|\omega\in A$ 或 $\omega\in B\}$记作 $C=A\cup B$.

类似地用 $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ 表示事件"$A_1,A_2,\cdots,A_n$ 几个事件中至少有一个发生".

$$\bigcup_{i=1}^{n} A_i = \{\omega|\omega \text{ 属于某一个 } A_i, i = 1,\cdots,n\}.$$

(3) **差**(difference) 记 C="A 发生而 B 不发生",称为事件 A 与 B 的差,$C=\{\omega|\omega\in A$,但 $\omega\notin B\}$,记作 $C=A\backslash B$(或 $A-B$).

注 并和交的运算可推广到无穷多个事件.

由上述三种基本的运算又可导出下列几种特殊的运算.

定义 1.1.7 设 Ω 为样本空间,A、B 为两个事件.

(1) **求余**(complementation)

$$\overline{A} = \Omega\backslash A$$

(2) **直和**(direct sum) 若 $AB=\varnothing$,则记

$$A + B = A \cup B$$

(3) **对称差**(symmetric difference)

$$A\triangle B = (A \cup B)\backslash(AB)$$

事件运算有下列的规律

$$\left.\begin{aligned}
&(1)\ \text{交换律}\quad A\cup B=B\cup A, AB=BA.\\
&(2)\ \text{结合律}\quad (A\cup B)\cup C=A\cup(B\cup C),\\
&\qquad\qquad\quad (AB)C=A(BC).\\
&(3)\ \text{分配律}\quad (A\cup B)C=(AC)\cup(BC),\\
&\qquad\qquad\quad (AB)\cup C=(A\cup C)(B\cup C).
\end{aligned}\right\}\qquad(1.1)$$

(4) 德摩根(De Morgan)律

$$\overline{(\bigcup_i A_i)} = \bigcap_i \overline{A}_i, \quad \overline{(\bigcap_i A_i)} = \bigcup_i \overline{A}_i. \tag{1.2}$$

(5) 直和分解律

$$\bigcup_{i=1}^{n} A_i = A_1 + (\overline{A}_1 A_2) + (\overline{A}_1 \overline{A}_2 A_3) + \cdots + (\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) \tag{1.3}$$

事件运算的优先顺序为:求余,对称差,交,差和并.

例 1.1.8 设事件 A,B,C 和 D 如例 1.1.1 中定义,则 $\overline{A}=\{\omega_1,\omega_4\}$, $BC=\{\omega_1,\omega_2\}=B$, $C-D=\{\omega_1\}$.

例 1.1.9 设 A,B,C 为三个事件,用它们之间的运算关系表示下列几个事件:

D="A,B,C 同时发生",

E="A,B,C 都不发生",

F="A,B,C 至少发生一个",

G="A,B,C 至多发生一个".

则

$D=ABC$.

$E=\overline{A}\,\overline{B}\,\overline{C}=\overline{A\cup B\cup C}$.

$F=A\cup B\cup C=A\,\overline{B}\,\overline{C}+\overline{A}B\,\overline{C}+\overline{A}\,\overline{B}\,C+AB\,\overline{C}+A\,\overline{B}C+\overline{A}BC+ABC$.

$G=\overline{A}\,\overline{B}\,\overline{C}+A\,\overline{B}\,\overline{C}+\overline{A}B\,\overline{C}+\overline{A}\,\overline{B}C$.

1.2 事件的概率

1.2.1 概率的定义及其频率解释

定义 1.2.1 设 Ω 为一个随机试验的样本空间,对 Ω 上的任一事件 A,规定一个实数与之对应记为 $P(A)$,满足下述三条基本性质,称为事件 A 发生的概率(probability):

(1) $0 \leqslant P(A) \leqslant 1$.

(2) $P(\Omega)=1, P(\varnothing)=0$.

(3) 若二事件 AB 互斥，即 $AB=\varnothing$，则

$$P(A \cup B) = P(A) + P(B).$$

上述三条基本规定是符合常识的.

例 1.2.2 设一个随机试验只有两个可能的结果，记为 ω_0 和 ω_1，则所有可能的事件只有 4 个：

$\Omega=\{\omega_0, \omega_1\}, \{\omega_0\}, \{\omega_1\}$ 和空集 $\varnothing$.

对每个事件定义一个概率如下：

$$P(\Omega) = 1, P(\{\omega_0\}) = 1 - p, P(\{\omega_1\}) = p, P(\varnothing) = 0.$$

其中 p 为介于 0 和 1 之间的一个正数. 容易验证，这样规定的概率满足定义 1.2.1 中要求的三条基本性质.

由此例看出，当 p 在 0 与 1 之间选不同的值时，事件的概率是不同的. 这说明，对具体事件赋概率时有一定的随意性. 为此需回答下面两个问题：

第一，概率的客观背景是什么？换言之，当人们说某一事件发生的概率为百分之几十时，其客观含义是什么？

第二，如何根据一定的试验结果来为事件赋恰当的概率或判断我们对某一事件所赋的概率是否恰当？

表 1.1 历史上掷硬币试验的记录

试验者	总试验次数 n	花面向上次数 m	频率 m/n
蒲丰	4040	2048	0.5070
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

人类由经验发现：当大量重复地进行随机试验时，随机事件的发生呈现出一定的规律性. 以抛掷硬币的试验为例，前人曾作过多

次试验，发现当抛掷次数很大时，花面和字面出现的次数相当接近，即大约各占总试验次数的 50%. 试验结果与人们按直觉判断的结论“出现花面和字面的可能性各为½”的理性分析相吻合.

除了掷硬币之外，人们还作了许多其它的试验，都显示了下述的客观规律性：

设某一随机试验可以在不改变试验条件的前提下互不干扰地重复进行任意多次，A 为一次试验所产生的任一个事件. 记 n 为试验次数，m_n 为在 n 次试验中 A 发生的次数，$f_n=m_n/n$ 为 A 在 n 次试验中发生的**频率**(frequency)，则当试验次数 n 很大且不断增大时，f_n 稳定地趋向于某一定值. 这种性质称为**频率的稳定性**.

由于频率的性质与定义 1.2.1 中所规定的三条基本性质基本相同，可以推出频率所趋向的定值也满足这三条基本性质. 因此这些定值可以解释为事件发生的概率. 这种用频率的“极限”来解释概率的观点被称为**概率的频率解释**或**统计解释**，频率的稳定性又可称为**经验大数定律**.

在统计物理学中，利用频率的稳定性建立概率模型来解释分子运动的规律. 在数理统计学中利用频率的稳定性来解答上面提出的第二个问题. 首先，当我们不知道某个事件的概率时，可以用频率去近似地估计它；其次，当我们假设某一事件发生的概率在一定范围内时，可以用频率去检验它.

当随机试验不能重复时，概率失去其频率解释的含义. 此时，概率还有其他解释. 围绕着对概率概念的不同解释已形成了三大学派：频率学派，Bayes 学派与信念(fiducal)学派.

1.2.2 概率的性质

除了定义 1.2.1 中所列的三条基本性质以外，概率还有一些重要的性质.

引理 1.2.3 概率的性质

(1) 单调性 设 $A\supset B$,则 $P(A)\geqslant P(B)$.

(2) 有限与可列可加性 设 $A_1,\cdots,A_n$ 为一列两两互斥的事件,即 $A_iA_j=\varnothing$,当 $i\neq j$,则

$$P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i).$$

$$P\left(\sum_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i).$$

(3) 连续性 设 $A_n\supset A_{n+1}$,$B_n\subset B_{n+1}$,$n=1,2,\cdots$,则

$$P\left(\bigcap_{n=1}^{\infty}A_n\right)=\lim_{n\to\infty}P(A_n).$$

$$P\left(\bigcup_{n=1}^{\infty}B_n\right)=\lim_{n\to\infty}P(B_n).$$

注 可列可加性与连续性并不能由定义 1.2.1 推出,因为该定义在数学上还不是十分严格的,读者可参阅 7.1 节.

引理 1.2.4 概率计算的常用公式

(1) $P(A-B)=P(A)-P(AB)$.

(2) $P(\overline{A})=1-P(A)$.

(3) $P(A\cup B)=P(A)+P(B)-P(AB)$.

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}[P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)].$$

(4) $P(A\triangle B)=P(A\cup B)-P(AB)$.

定义 1.2.5 设$\{A_n,n=1,2,\cdots\}$为一组有限或可列无穷多个事件,两两不相交且 $\sum_{n}A_n=\Omega$,则称事件族$\{A_n,n=1,2,\cdots\}$为样本空间 Ω 的一个**完备事件族**.又若对任一事件 B 有 $BA_n=A_n$ 或$\varnothing$,$n=1,2,\cdots$,则称$\{A_n,n=1,2,\cdots\}$为**基本事件族**.

定理 1.2.6 若$\{A_n,n=1,2,\cdots\}$为一完备事件族,则

$\sum_n P(A_n)=1$，且对任一事件 B 有

$$P(B)=\sum_n P(A_nB).$$

又若 $\{A_n, n=1,2,\cdots\}$ 为一基本事件族，则

$$P(B)=\sum_{A_n \subset B} P(A_n).$$

例 1.2.7 设一袋中有 N 个红球和 M 个白球. 从中无放回地抽取 n 个球. 记 N 个红球为 $R_1,R_2,\cdots,R_N$，M 个白球为 $W_1,W_2,\cdots,W_M$. 则每个样本点为 n 个 R_i 和 W_j 的一个排列. 假设抽球记录只考虑抽到几个什么样的红球和几个什么样的白球，而不考虑其抽到的先后顺序，则每个基本事件为 n 个 R_i 和 W_j 的组合. 由于每个基本事件中有相同多的样本点($n!$ 个)，因此，在古典概率模型中(见节 1.3)，每个基本事件有相同的概率 $1\Big/\binom{N+M}{n}$. 于是对事件 B_k="抽到 k 个红球"($k\leqslant n$)，其中包含了 $\binom{N}{k}\binom{M}{n-k}$ 个基本事件，故 $P(B_k)=\binom{N}{k}\binom{M}{n-k}\Big/\binom{N+M}{n}$.

1.2.3 概率分布

定义 1.2.8 设 Ω 为样本空间，对 Ω 上的所有事件依照定义 1.2.1 的规定确定了概率，则 $\{P(A) \mid A$ 为 Ω 上的事件$\}$ 构成样本空间 Ω 上的一个**概率分布**(probability distribution).

若 Ω 上存在一族有限或可列多个基本事件，且对每个基本事件规定了发生的概率，则由定理 1.2.6 可知任一事件的概率就由这些基本事件的概率所决定.

例 1.2.9 设样本空间有有限个样本点，$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，$n$ 为样本点总数. 则事件 $\{\omega_i\}, i=1,2,\cdots,n$(即由单个样本点构成的事件)为一基本事件族，若规定 $P(\{\omega_i\})=\dfrac{1}{n}, i=1,\cdots,n$. 则由此确定的概率分布称为**等概率分布**或古典概率分布.

例 1.2.10 设样本空间 $\Omega=\{0,1,2,\cdots\}$，对基本事件族 $\{k\}$，$k=0,1,2,\cdots$，定义概率为

$$P(\{k\})=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots,$$

则确定了 Ω 上的一个概率分布，称为泊松(Poisson)分布(详见节 6.2).

然而在很多场合并不存在一个有限或可列的基本事件族，此时，概率分布的表达形式比较复杂. 例如：Ω 为实数空间 R^1 或 R^2 上的一个区间. 这时，事件的概率就要用积分的形式来表示.

例 1.2.11 设 $\Omega=\mathbf{R}^1$，Ω 上的任一事件可由有限或可列无穷多个区间的并和交的运算得到，对任一区间 $(a,b]$，定义

$$P((a,b])=\int_a^b \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx.$$

这样确定的概率分布称为**标准正态分布**(standard normal distribution)(详见 6.4).

1.3 古典概率

定义 1.3.1 设样本空间 Ω 由有限个样本点构成，若每个样本点作为一个基本事件有相同的发生概率，如此确定的概率分布称为**古典概率分布**(classical probability distribution).

推论 1.3.2 设 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，在 Ω 上定义了一个古典概率分布，则有 $P(\{\omega_i\})=\frac{1}{n}$，$i=1,\cdots,n$，且对任一事件 A 有

$$P(A)=\frac{1}{n}(A\text{ 中样本点数}).$$

古典概率分布是一种理论上简单而又有广泛实用价值的概率模型. 使用古典概率模型的前提是样本点具有某种对称性或经受住频率检验. 例如掷一枚硬币，观察其向上的面，如果可以认为硬

币的形状是标准的，质地是均匀的，就没有理由认为某一面向上的概率比另一面更大些. 频率检验也证明了这一点(见表 1.1).

例 1.3.3 在一个由 N 个**个体**(individual)组成的**总体**(population)中逐次抽取 n 个个体，称为**抽样**(sampling). 假如在抽样过程中每抽到一个个体后做记录，再将此个体放回，抽下一个. 这样的抽样称为**有放回抽样**(sampling with replacement). 在有放回抽样中，每个结果是 n 个个体(允许重复)的一个排列. 样本点总数为 N^n. 因此在古典概率模型中每个样本点发生的概率为 $1/N^n$.

例 1.3.4 设在一袋内有 N 个白球和 M 个黑球，从中有放回地随机抽取 n 个球，则抽到 k 个白球的概率为

$$P_k = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n.$$

其中

$$p=\frac{N}{N+M}, \quad q=\frac{M}{N+M}.$$

分析如下：由于样本点的总数为 $(N+M)^n$，因此每个样本点发生的概率为 $1/(N+M)$. 记 A_k 为事件"抽到 k 个白球"，则 A_k 中的样本点可按以下两步分类. 首先考虑这 k 个白球是在哪几次抽取中获得的可分成 $\binom{n}{k}$ 个类. 其次，从 N 个白球中取 k 个，在有放回抽样场合有 N^k 种可能性，最后从 M 个黑球中取 $n-k$ 个有 M^{n-k} 种可能性. 于是 A_k 中样本点数为 $\binom{n}{k} N^k M^{n-k}$，再由推论 1.3.2 即得 P_k 的表达式.

利用牛顿二项公式容易验证 $\sum\limits_{k=0}^{n} P_k = 1$. 因此 $\{P_k\} k=0,1,2,\cdots,n\}$，在样本空间 $\Omega=\{0,1,\cdots,n\}$ 上确定了一个概率分布，称为**二项分布**(binomial distribution)(详见 6.1).

例 1.3.5 在一个由 N 个个体构成的总体中随机抽取 n 个个体. 假定抽样过程是每次抽取一个个体，记录后不再放回总体中，再抽下一个，直到抽到 n 个个体为止($n \leqslant N$). 这种抽样称为**无放**

回抽样(sampling without replacement). 在无放回抽样场合,样本点总数为 $N(N-1)\cdots(N-n+1)$. 若不考虑抽样的顺序,则有 $\binom{N}{n}$ 个基本事件,每个包含 $n!$ 个样本点.

例 1.3.6 设一袋中有 N 个白球, M 个黑球,从中无放回地抽取 n 个,则抽到 k 个白球的概率为

$$h_k=\binom{N}{k}\binom{M}{n-k}\Big/\binom{M+N}{n},k=0,1,2,\cdots,n.$$

容易验证, $\sum_{k=0}^{n}h_k=1$. 因此 $\{h_k\mid k=0,1,\cdots,n\}$ 在样本空间 $\Omega=\{0,1,\cdots,n\}$ 上确定了一个概率分布,称为**超几何分布**(hypergeometric distribution)(详见 6.3.7).

h_k 的计算方法如下:在无放回抽样中,不考虑顺序,则共有 $\binom{N+M}{n}$ 个基本事件,每个基本事件有相同个数的样本点($n!$ 个). 因此,每个基本事件的概率为 $\binom{N+M}{n}^{-1}$. 记 B_k 为抽到 k 个白球的事件,则 B_k 中的基本事件可按下列两步分类. 首先考虑包含哪 k 个白球,可分 $\binom{N}{k}$ 类,其次再考虑包含哪 $n-k$ 个黑球,可分 $\binom{M}{n-k}$ 类. 故 B_k 中共包含 $\binom{N}{k}\binom{M}{n-k}$ 个基本事件. 因此由定理 1.2.6 知 $h_k=\binom{N}{k}\binom{M}{n-k}\Big/\binom{N+M}{n}$.

在物理学中研究质点在相空间中散布的情况,要建立恰当的概率模型. 设有 r 个质点,而相空间被划分成 n 个小区域,则 r 个质点在 n 个小区域中的散布情形,可以用 r 个球投入 n 个盒的模型来形象地描述. 有三种不同的基于古典概率的物理模型.

例 1.3.7 麦克斯威尔-波尔茨曼(Maxwell-Boltzmann)模型 设 r 个球是可分辨的,且每个盒子可容纳的球数没有限制. 则 r 个球在 n 个盒中的可能的散布状态共有 n^r 种,每种可能的散布状态发生的概率为 n^{-r}.

例 1.3.8 波司-爱因斯坦(Bose-Einstein)模型 设 r 个球不可分辨;每个盒子可容纳的球数也没有限制. 则 r 个球在 n 个盒中的散布状态仅由球分配到每个盒中的个数来确定,共有 $\binom{n+r-1}{n-1}$ 种可能的状态,每种可能的状态发生的概率为 $\binom{n+r-1}{n-1}^{-1}$.

例 1.3.9 费米-狄拉克(Fermi-Dirac)模型 设 r 个球不可分辨,且每个盒中至多只能容纳一个球. 则有 $\binom{n}{r}$ 种散布状态,每种状态发生的概率为 $\binom{n}{r}^{-1}$.

上述三种模型适合于描述不同的物理粒子的运动状态.

例 1.3.10 数字 $1,2,\cdots,N$ 的一个随机排列中,若数字 k 恰排在第 k 个位置上则称为一个**相合**(congrence). 在一个排列中相合的个数可能为:$0,1,2,\cdots,N-2,N$. 由于排列的总数为 $N!$,因此每个排列发生的概率为 $(N!)^{-1}$. 用 A_i 记事件"在第 i 个位置上发生相合",则

$$P(A_i) = (N-1)!/N! = \frac{1}{N}, i = 1,2,\cdots,N.$$

$$P(A_iA_j) = (N-2)!/N! = \frac{1}{N(N-1)}, i < j.$$

$$P(A_iA_jA_k) = (N-3)!/N! = \frac{1}{N(N-1)(N-2)}, \quad i < j < k.$$

……

用 C 记事件"至少有一个相合发生",则由公式(1.2.4)有

$$\begin{aligned} P(C) &= P\left(\bigcup_{i=1}^{N} A_i\right) \\ &= \binom{N}{1}\frac{1}{N} - \binom{N}{2}\frac{1}{N(N-1)} + \binom{N}{3}\frac{1}{N(N-1)(N-2)} \\ &\quad - \cdots + (-1)^{N-3}\frac{\binom{N}{N-2}}{N(N-1)\cdots 3} + (-1)^{N-1}\frac{1}{N!} \end{aligned}$$

$$=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{N-3}\frac{1}{(N-2)!}$$

$$+(-1)^{N-1}\frac{1}{N!}$$

$$\approx e^{-1}.$$

例 1.3.11 设 n 个字母 A 和 m 个字母 B 排成一列，则这一列字母可分成若干个小段，每段内的字母相同，而相邻段的字母不同，这样的一个小段称为一个**游程**(run)或**连贯**. 如，在排列 $AAABAABBA$ 中有 5 个游程，即 3 个 A 游程 AAA，AA 及 A，2 个 B 游程 B 及 BB. n 个 A 和 m 个 B 的排列共有 $\binom{n+m}{n}$ 种可区分的结果. 而一个排列中有 k 个 A 游程相当于把 n 个不可分辨的球投入 k 个盒中且没有一个盒是空的，共有 $\binom{n-1}{k-1}$ 种可能的状态. 同理，一个排列中有 l 个 B 游程的可能的状态为 $\binom{m-1}{l-1}$ 种，但 $|l-k|\leqslant 1$. 据此可算出：

$$P(k\text{ 个 }A\text{ 游程},k-1\text{ 个 }B\text{ 游程})=\binom{n-1}{k-1}\binom{m-1}{k-2}\Big/\binom{n+m}{n}.$$

$$P(k\text{ 个 }A\text{ 游程},k\text{ 个 }B\text{ 游程})=2\binom{n-1}{k-1}\binom{m-1}{k-1}\Big/\binom{n+m}{n}.$$

$$P(k\text{ 个 }A\text{ 游程},k+1\text{ 个 }B\text{ 游程})=\binom{n-1}{k-1}\binom{m-1}{k}\Big/\binom{n+m}{n}.$$

由以上三式可推得

$$P(k\text{ 个 }A\text{ 游程})=\binom{n-1}{k-1}\binom{m+1}{k}\Big/\binom{n+m}{n}.$$

1.4 几何概率

作为等概率模型，古典概率只适用于有限样本空间. 但利用古典概率等于事件中样本点的个数与样本点总数之比的性质，可推广到一类定义在无限样本空间上的等概率模型.

设样本空间 Ω 为一维直线(或二维平面，或三维空间)上的有

界区间(或有界平面区域,或有界空间区域),则对任一事件 A,可定义其概率如下:

$$P(A)=\frac{A\text{ 的长度(面积,体积)}}{\Omega\text{ 的长度(面积,体积)}}.$$

这样定义的概率满足定义 1.2.1 中规定的三条性质.由此产生的概率模型称为**几何概率模型**(geometric probability model).

例 1.4.1 设一个球在一个半径为 1 的圆盘上作随机滚动后停止.则此球停在圆盘上任一区域 B 中的概率为

$$P=\frac{B\text{ 的面积}}{\text{圆盘的面积}}=\frac{1}{\pi}(B\text{ 的面积}).$$

这是一个几何概率模型.

例 1.4.2 蒲丰试验(Buffon Experiment) 法国科学家蒲丰曾作过如下的试验:在平面上画一些相隔距离为 a 的平行线,向此平面上随意地投一根长度为 $l(l<a)$的针,求此针与任一平行线相交的概率.

解 以 x 表示针的中点到最近一条平行线的距离,φ 表示针与平行线的夹角.针与平行线的关系由图 1.1(a)所示.则样本空间 $\Omega=\{(x,\varphi)\mid 0\leqslant x\leqslant\frac{a}{2},0\leqslant\varphi\leqslant\pi\}$.针与平行线相交的条件为 $x\leqslant\frac{l}{2}\sin\varphi$,满足此条件的区域记为 A,在图 1.1(b)中用阴影部分

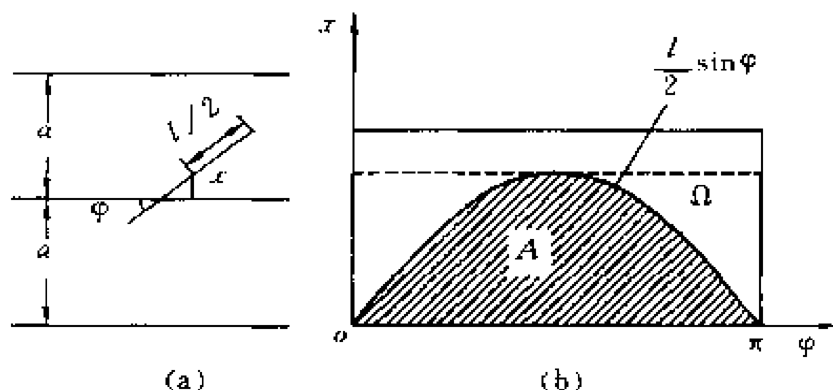


图 1.1

表示出.这是一个几何概率模型

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{\frac{1}{2}\int_0^{\pi} l\sin\varphi\,\mathrm{d}\varphi}{\frac{1}{2}a\pi}=\frac{2l}{\pi a}.$$

将上式变形后有 $\pi=2l/(aP(A))$. 蒲丰(Buffon)作了大量投针试验,用频率代替概率来近似计算 π 的值.这是随机模拟的雏形.

例 1.4.3　贝特朗奇论(Bertrand paradox)　在半径为 1 的圆内随机地取一条弦,问其长超过圆内接等边三角形边长的概率为多少?

这问题有许多不同的解,下面列出其中三种:

解 1　任何弦交圆周两点.不失一般性,先固定其中一点,则弦的另一端只有位于此固定点正对的$\frac{1}{3}$圆周长的弧内才满足要求,故所求概率为$\frac{1}{3}$(见图 1.2(a)).

解 2　弦长只与它到圆心的距离有关.因此满足要求的弦和与之垂直的直径的交点到圆心的距离必须大于$\frac{1}{2}$,而直径长为 2,故所求概率为$\frac{1}{2}$(见图 1.2(b)).

解 3　弦的中心点必须位于半径为$\frac{1}{2}$的同心圆之内才满足要求.而此圆面积为大圆面积的$\frac{1}{4}$,故所求概率为$\frac{1}{4}$(见图1.2(c)).

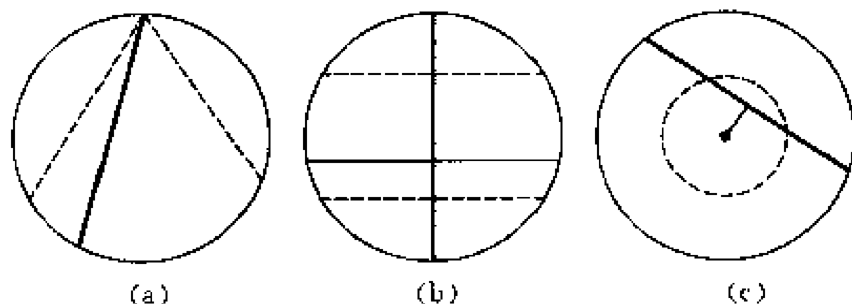


图　1.2

为什么对“同一个”问题会产生出三种各自有理而又互不相同的解答呢?这里面的关键问题在于例 1.4.3 中所说的“随机地取一条弦”一句话中,“随机”二字的含义不明确.而上述三个答案各自给出了特定的随机性解释.因此在各自的“随机”的概念下,三个解答都是正确的.

2 条件概率与事件的独立性

2.1 条件概率

2.1.1 条件概率的定义

定义 2.1.1 设 A,B 为事件且 $P(B)>0$，称 $P(A|B)=P(AB)/P(B)$为在事件 B 已发生的条件下，事件 A 发生的**条件概率**(conditional probability).

简称 $P(A|B)$为给定 B 时 A 发生的条件概率.

条件概率的概念是基于一种缩小样本空间的思想：即已知 B 发生，则只考虑属于 B 的那些样本点. 这由 $P(B|B)=1$ 及当 $AB=\varnothing$时，$P(A|B)=0$ 可以看出. 另一方面当 B 已发生时则 A 也发生意味着 A 与 B 同时发生，因此 $P(A|B)$与 $P(AB)$成正比，而比例因子为 $1/P(B)$.

下面用条件频率和古典条件概率为例来解释条件概率的概念.

设 A,B 为随机试验 E 所产生的两个事件. 将 E 重复且互不影响地作 n 次，用 m_A 和 m_B 分别记 A,B 发生的次数，f_A,f_B 分别为 A,B 发生的频率. 则 $f_B=m_B/n$，而 $f_{AB}=m_{AB}/n$，其中 m_{AB}和 f_{AB} 分别为事件 AB 发生的次数和频率. 由经验大数定律(见 1.2.1)可以近似地用 $f_{AB}/f_B=m_{AB}/m_B$ 去逼近 $P(A|B)$. 而 m_{AB}/m_B 是在只考虑 B 发生的那些试验中 A 发生的频率.

再考虑古典概率模型. 用 N,N_B 和 N_{AB}分别记样本点总数，B 中样本点数和 AB 中样本点数. 则对古典概率模型有

$$P(A|B)=P(AB)/P(B)=\frac{N_{AB}}{N}\bigg/\frac{N_B}{N}=N_{AB}/N_B.$$

故 $P(A|B)$定义了在缩小的样本空间 B 上的古典概率.

2.1.2 条件概率的性质

如上节所述,条件概率实质上是在缩小的样本空间上的概率分布;另一方面,它在形式上仍然在原来的样本空间上定义了一个概率分布:

$$\left.\begin{array}{l}(1)\ 0\leqslant P(A|B)\leqslant 1.\\(2)\ P(\Omega|B)=1,P(\varnothing|B)=0.\\(3)\ 若\ A_1A_2=\varnothing,则\ P(A_1+A_2|B)=P(A_1|B)+P(A_2|B).\end{array}\right\}\tag{2.1}$$

乘法公式:

$$\left.\begin{array}{l}(1)\ P(AB)=P(B)P(A|B).\\(2)\ P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots\\\qquad\qquad P(A_n|A_1A_2\cdots A_{n-1}).\end{array}\right\}\tag{2.2}$$

定理 2.1.2 全概率公式: 设 $A_1,A_2,\cdots,A_n$ 互不相交,$\sum_{i=1}^{n}A_i=\Omega$,且 $P(A_i)>0,i=1,\cdots,n$,则对任一事件 A 有

$$P(A)=\sum_{i=1}^{n}P(A_i)P(A|A_i).$$

例 2.1.3 从一个含有 N 个个体的总体中进行无放回抽样.设 a,b 为两个个体,则不论抽多少次,第一次抽到 a 且第二次抽到 b 的概率为$\frac{1}{N(N-1)}$.这是因为:第一次抽到 a 的概率为$\frac{1}{N}$,而在此条件下,第二次抽到 b 的概率为$\frac{1}{N-1}$(因为只能在除 a 以外的 $N-1$ 个个体中抽取).因此用乘法公式(2.2)即得结论.

例 2.1.4 从一个总体 Ω(例如一群人)中随机抽取一个具有

某种特征的个体(例如左撇子),可以采用分层抽样的方法:将此总体分成若干互不相交的子体(例如按民族区分). 分别记为 $\Omega_1, \Omega_2, \cdots, \Omega_m$. 先以一定的概率分布从这些子体中进行抽取,记抽到 Ω_i 的概率为 $p_i, i=1,\cdots,m$, $\sum_{i=1}^{m} p_i = 1$. 再从抽到的子体中依一定的概率分布抽样. 设从 Ω_i 中抽到具有指定特征的个体的概率为 q_i, $i=1,\cdots,m$. 则由全概率公式 2.1.2 知,总的概率为 $p = \sum_{i=1}^{m} p_i q_i$.

例 2.1.5 波利亚(Polyá)罐子模型 设一个罐子中有 a 个黑球,b 个红球. 现从中进行有放回抽样. 每当抽出一个球时就向罐中再放进 c 个同颜色的球,这样反复抽取 n 次. 容易由乘法公式算出,不论抽出球的顺序如何,抽到 m 个黑球,$n-m$ 个黑球的概率为

$$
\begin{aligned}
P =& a(a+c)(a+2c)\cdots(a+(m-1)c)\cdot b\cdot(b+c)(b+2c) \\
& \cdots(b+(n-m-1)c)/[(a+b)(a+b+c)(a+b+2c) \\
& \cdots(a+b+(n-1)c)]
\end{aligned}
$$

这个模型是首先由波利亚(Polyá)提出,它已作为研究传染病现象的一个简化模型.

例 2.1.6 设 n 个人从 N 张签中依次随意抽一张,则容易由乘法公式算出,每个人抽到某张指定签的概率为 $1/N$. 这结论称为抽签原理.

2.1.3 贝叶斯公式

定理 2.1.7 贝叶斯(Bayes)公式 设 $A, B_1, \cdots, B_n$ 为一些事件,$P(A)>0$, $B_1,\cdots,B_n$ 互不相交,$P(B_i)>0, i=1,\cdots,n$,且 $\sum_{i=1}^{n} P(B_i) = 1$,则对 $k=1,\cdots,n$,

$$P(B_k|A)=\frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n}P(B_i)P(A|B_i)}.$$

贝叶斯公式容易由条件概率的定义,乘法公式和全概率公式得到. 在贝叶斯公式中,$P(B_i)$,$i=1,\cdots,n$ 称为**先验概率**,而 $P(B_i|A)$,$i=1,\cdots,n$ 称为**后验概率**. 贝叶斯公式的含义可解释如下:$B_1,B_2,\cdots,B_n$ 为 n 个互不相容的"原因",而 A 为某种"结果". 在实际问题中,"原因"发生的概率($P(B_i)$)和已知由某种"原因"产生结果的概率($P(A|B_i)$)都是可以事先估计的,则我们可以用贝叶斯公式反过来计算已知"结果"而它是某个"原因"产生的条件概率($P(B_k|A)$). 当某个 $P(B_k|A)$比较大时,则一观察到 A 就首先考虑是否由 B_k 引起的;另一方面,即使 $P(B_k|A)$的值不大,但它与$P(B_k)$相比却大大增加了,这现象说明 B_k 与 A 有很紧密的联系,因而须加以充分的重视.

例 2.1.8 用血清甲胎蛋白法诊断肝癌具有如下效果:对肝癌患者检验阳性反应达 95%,而对非肝癌患者检验也有 1%的阳性反应. 又根据统计资料得知人群中肝癌的发病率为万分之四. 现用此方法诊断某病人,检验结果为阳性,问此人确实患肝癌的概率为多大?

这是一个求条件概率的问题,用 A 记事件"化验呈阳性",B 记事件"病人患肝癌",要求 $P(B|A)$. 由题意知 $P(A|B)=0.95$,$P(A|\overline{B})=0.01$,$P(B)=0.0004$. 用贝叶斯公式求得

$$P(B|A)=\frac{P(B)P(A|B)}{P(B)P(A|B)+P(\overline{B})P(A|\overline{B})}=0.037.$$

可见,化验呈阳性而实际患肝癌的概率并不高. 要提高此化验法的准确性关键是要降低 $P(A|\overline{B})$. 尽管如此,经检验反应呈阳性时,病人患肝癌的机会比未检验前提高了近 100 倍.

2.2 事件的独立性

定义 2.2.1 设 A,B 为两个事件,满足

$$P(AB)=P(A)\cdot P(B),$$

则称事件 A 与事件 B **相互独立**(mutually independent),简称事件 A 与 B 独立.

引理 2.2.2 (1) 若 $P(A)=0$ 或 1,则 A 与任一事件独立.

(2) 若 A 与 B 独立且 $P(B)>0$,则

$$P(A|B)=P(A).$$

(3) 若 A 与 B 独立,则 A 与 $\overline{B}$,$\overline{A}$ 与 B,$\overline{A}$ 与 $\overline{B}$ 都是相互独立的事件对.

定义 2.2.3 设 $A_1,A_2,\cdots,A_n$ 为 n 个事件,满足下述条件:

$$P(A_iA_j)=P(A_i)P(A_j),\quad 1\leqslant i<j\leqslant n,$$

$$P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k),\quad 1\leqslant i<j<k\leqslant n,$$

$$\cdots\cdots,$$

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n),$$

则称事件 $A_1,A_2,\cdots,A_n$ 相互独立.

在定义 2.2.3 中共有 $\binom{n}{2}+\binom{n}{3}+\cdots+\binom{n}{n}=2^n-n-1$ 个约束条件,定义了 n 个事件相互独立的条件.

推论 2.2.4 若 n 个事件相互独立,则对 $m<n$,其中任意 m 个事件也是相互独立的.

推论 2.2.4 的逆命题不成立.

例 2.2.5 设有 4 张卡片,其中 3 张上分别记有字母 A 和 B,B 和 C 及 A 和 C,第 4 张卡片是空白的.从中随机地抽一张,就用 A,(B 和 C)分别记事件"抽到的卡片上有字母 A,(B 和 C)",则显然有

$$P(A)=P(B)=P(C)=\frac{1}{2},$$

$$P(AB)=P(AC)=P(BC)=\frac{1}{4}.$$

但 $P(ABC)=0$. 因此 A,B,C 三个事件中任意两个相互独立，但这三个事件并不相互独立.

推论 2.2.6 设 n 个事件相互独立，则对 $0\leqslant m\leqslant n$，其中任意 m 个事件与其余 $n-m$ 个事件的对立事件构成 n 个相互独立的事件.

在实际问题中，判断事件是否独立，不能仅从数学定义出发，而要根据问题的实际的背景来判断. 例如在一电路中有若干元件，每个元件都有可能损坏，因而有相应的失效概率. 如果元件失效与否并不互相影响，则可以认为这些元件失效与否是相互独立的.

事实上在实际问题中，事件的独立性来源于试验的独立性(见 2.3).

2.3 试验的独立性

定义 2.3.1 设 $E_1,E_2,\cdots,E_n$ 为 n 个随机试验且假定每个试验产生的可能结果及结果的概率不受其它试验结果的影响，则称这 n 个试验是**相互独立的试验**.

直观地讲，因为这 n 个试验是相互独立的，则设 A_i 为由 E_i 产生的事件，$i=1,\cdots,n$，$A_1,A_2,\cdots,A_n$ 即为相互独立的事件. 这里要回答一个疑问，即 $A_1,\cdots,A_n$ 是属于由不同试验产生的样本空间. 而我们在第 1.1 节中强调过，当讨论事件之间的关系时总是考虑在同一样本空间上的事件. 这个矛盾如何解决? 要解答这个疑问，就要引进复合试验的概念.

定义 2.3.2 设随机试验 E 是由 n 个试验 $E_1,E_2,\cdots,E_n$ 按顺

序进行所构成的，则称 E 是由 $E_1,E_2,\cdots,E_n$ **复合**而成，记为 $E=E_1\times E_2\times\cdots\times E_n$. 记 Ω_i 为由 E_i 生成的样本空间，$i=1,\cdots,n$，则 E 的样本空间 $\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n$，$\Omega$ 中的样本点 ω 可表为 n 元组 $(\omega_1,\omega_2,\cdots,\omega_n)$，其中 ω_i 为 Ω_i 中的样本点. Ω **称为乘积样本空间**.

有了试验的复合与乘积样本空间的概念之后，对 E_i 产生的任一事件 A_i 便可从两个等价的角度去理解. 当我们只考虑试验 E_i 及其结果时，就把 A_i 仅看成是 Ω_i 的子集并计算其概率. 当考虑复合试验 $E=\overset{n}{\underset{i=1}{X}}E_i$ 时，A_i 便被看作是 Ω 中的子集，其元素为 $\{(\omega_1,\cdots,\omega_i,\cdots,\omega_n)\mid\omega_i\in A_i,\omega_j\in\Omega_j,j\neq i\}$. 通常把 A_i 这样的事件称为 Ω 中的**柱形集**.

这样，当我们说由 E_i 产生的事件 $A_i,i=1,\cdots,n$ 是**相互独立**时，我们是把 $A_1,A_2,\cdots,A_n$ 看成乘积样本空间 Ω 上的子集.

定义 2.3.3 设 E 为一随机试验，将 E 独立地重复 n 次得到的复合试验便称为 **n 重试验**.

定义 2.3.4 伯努利试验(Bernowlli trials) 设随机试验 E 只产生两个对立的结果，分别称为“成功”与“失败”，记 p 为“成功”的概率，则 E 称为“成功”概率为 p 的伯努利试验. 若将此试验独立地重复进行 n 次，便称为 **n 重伯努利试验**.

例 2.3.5 设从一批产品中检验次品，在其中进行有放回抽样 n 次，抽到次品称为“成功”，抽到正品称为“失败”，这就是 n 重伯努利试验.

例 2.3.6 设 n 个灯泡是从一大批灯泡中随机抽出来的，又假定每个灯泡可以点燃一千小时以上的概率是相同的，记为 p. 将这 n 个灯泡进行编号，然后同时点燃. 对每个灯泡来说，点燃一千小时以上称为“成功”；否则称为“失败”. 这也是一个 n 重伯努利试验.

在一个“成功”概率为 p 的 n 重伯努利试验中，有 k 次“成功”

($n-k$ 次"失败")的概率为

$$P_k = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0,1,\cdots,n. \qquad (2.3)$$

显然$\{P_k, k=0,1,\cdots,n\}$构成一个二项概率分布(见 1.3.6).

公式(2.3)是这样得到的:利用试验之间的独立性可知,一个包含 k 次"成功",$n-k$ 次"失败"的试验结果序列发生的概率为 $p^k(1-p)^{n-k}$,而这样的试验结果序列有$\binom{n}{k}$个.

在 n 重伯努利试验中试验次数 n 是事先给定的. 另一种进行伯努利试验的方法是:不固定试验次数 n,当试验进行到出现一次"成功"时便停止,记 P_k 为在第 k 次出现首次"成功"的概率,有

$$P_k = (1-p)^{k-1} p \quad k = 1,2,\cdots. \qquad (2.4)$$

$\{P_k, k=1,2,\cdots\}$构成一个几何分布(见 6.1.3).

如果将伯努利试验序列进行到出现 r 次"成功"便结束,记 P_k 为当进行到第 k 次试验时出现第 r 次"成功"的概率,则

$$P_k = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad k = r, r+1, \cdots. \qquad (2.5)$$

$\{P_k, k=r, r+1, \cdots\}$构成一个帕斯卡(Pascal)分布(见 6.1.10).

3 随机变量及其分布

3.1 随机变量与随机向量

3.1.1 随机变量的定义

定义 3.1.1 设 Ω 为某一试验产生的样本空间，X 为定义在 Ω 上的实函数，即对任一样本点 $\omega\in\Omega$，$X(\omega)$ 为一实数，则称 X 为一个**随机变量**(random variable).

注 严格来说，X 应为可测函数(见定义 7.2.4)，因此上述定义从数学的角度看不够严格，但对实际应用不会产生什么大的影响.

对随机试验所产生的结果(样本点)，人们通常所关心的是与之有关的一个数量指标. 例如，从一批产品中抽出 n 个样品，统计其中次品的个数 X；从一批灯泡中抽出一个灯泡，检测它的寿命 Y；从一群儿童中抽出一个，化验他(她)的血色素 Z，等等. 由于试验的结果是随机的，因而与之相关联的数量指标(X，Y，Z，等)的取值也是随机的，这些数量指标就是随机变量.

由于随机变量的取值是随机的，所以我们关心其取某值或在某个范围内取值的概率是多少.

定义 3.1.2 设 P 为样本空间 Ω 上的一个概率分布，X 为 Ω 上的一个随机变量，对任一实数集合 A，定义

$$P(X\in A)=P(\{\omega\mid\omega\in\Omega,X(\omega)\in A\}),$$

称为随机变量 X 属于集合 A 的概率.

注 严格来说，上述定义中的集合 A 应为可测集(见 7.1.5). 但因常见的集合大都是可测集，因此这样的定义并不妨碍我们

使用.

定义 3.1.2 中等式的含义是：事件"$X\in A$"的概率由样本空间 Ω 上一个事件的概率来确定，这事件是由所有满足 $X(\omega)\in A$ 的样本点 ω 构成.

例 3.1.3 将一枚均匀的硬币掷 n 次，每次记下花面向上(H)或字面向上(W)，则样本点为由 H 和 W 组成的一个 n 元排列. 记 X 为正面向上的次数($X(\omega)=$"ω 中 H 的个数")，则 X 为随机变量. $P(X=k)=P(\{\omega:X(\omega)=k\})=\binom{n}{k}2^{-n}$，$k=0,1,2,\cdots,n$.

3.1.2 随机变量的类型

常用的随机变量按其取可能值的情况分成三类.

定义 3.1.4 若随机变量 X 只可能在有限或可列无穷多个(实数)点上取值，则称 X 为**离散型随机变量**.

记 X 的所有可能值为$\{x_k,k=1,2,\cdots\}$，记 $p_k=P(X=x_k)$，$k=1,2,\cdots$，则$\{(x_k,p_k),k=1,2,\cdots\}$，称为 X 的**分布列**.

分布列决定了离散型随机变量的概率分布. 在实用中常用下述三种方式表示分布列.

(1) 公式法　分布列用公式表示，例如

$$P(X=k)=p_k=\frac{\lambda^k}{k!}e^{-\lambda}\quad k=0,1,2,\cdots,\lambda\text{ 为正常数}.$$

其中 $p_k=f(k)$，$k=1,2,\cdots$.

(2) 列表法　将分布列用表格方式表示，例如

x	0	1	2	3
p	0.2	0.3	0.3	0.2

(3) 图示法　将随机变量的可能值标示在横坐标轴上，用垂

直线表示取该可能值的概率. 例如上面分布列可如图 3.1 所示.

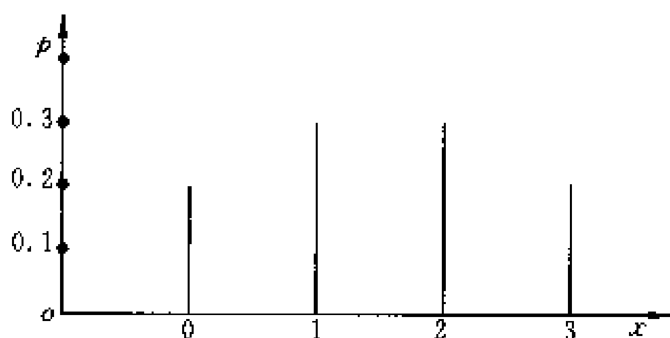


图 3.1

分布列具有下述性质.

推论 3.1.5 设$\{(x_k,p_k),k=1,2,\cdots\}$为随机变量的分布列,则有 $p_k\geqslant 0,k=1,2,\cdots$,且 $\sum_k p_k=1$.

定义 3.1.6 设随机变量 X 在一个或多个非退化的实数区间上可以连续取值,且存在一个非负的实函数 $f(x)$,使得对任一区间(a,b),有 $P(x\in(a,b))=\int_a^b f(x)\mathrm{d}x$,则称 X 为**连续型随机变量**,$f(x)$称为 X 的**概率密度函数**(probability density function),简称为**密度**.

如同分布列决定了离散型随机变量的概率分布一样,密度决定了连续型随机变量的概率分布.

推论 3.1.7 设 $f(x)$为随机变量的概率密度函数,则有 $f(x)\geqslant 0$,且$\int_{-\infty}^{\infty}f(x)\mathrm{d}x=1$.

密度 $f(x)$的表示方法有两种:解析表示法和图示法.

例 3.1.8 设随机变量 X 有密度

$$f(x)=\begin{cases}0, & \text{当 } x<0,\\ \lambda e^{-\lambda x} & \text{当 } x\geqslant 0.\end{cases}\quad \lambda \text{ 为正常数.}$$

容易验证 $f(x)$满足推论 3.1.7 中的结论,其图形如图 3.2.

图中曲线为密度 $f(x)$的图形,阴影中的面积表示

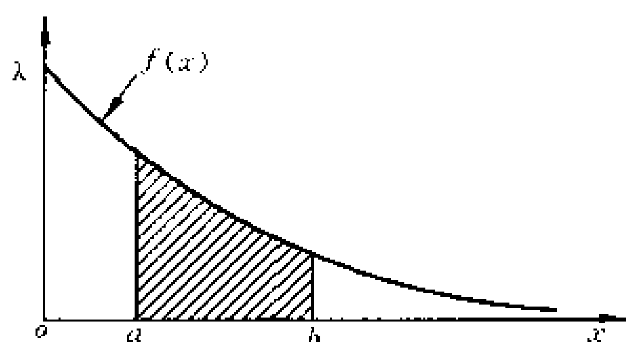


图 3.2

$$P(X \in (a,b)).$$

除了上述两大类随机变量之外，有时（特别在可靠性统计中）会遇到一种混合型随机变量．即随机变量在某些离散点上取大于零的概率，而在其它地方的概率可由积分表示．

例 3.1.9 设灯泡寿命有密度 $f(t)=e^{-t}, 0<t$，作试验如下：从一批灯泡中抽一个点燃，当其寿命在 1000 小时以内时，记 $X=$ 灯泡寿命，当其寿命超过 1000 小时后，记 $x=1000$．于是 x 即为一个混合型随机变量，它在 $(0,1000)$ 上有密度 $f(t)=e^{-t}$，而在 1000 上有概率 $1-\int_0^{1000} e^{-t}dt = e^{-1000}$．

离散型随机变量与连续型随机变量在数学性质上有很大的差别，但在实际使用时不可绝对化．设某个离散型随机变量在一个区间内密集地取值，可以近似地看成在此区间内连续取值，因此可以用一个连续型随机变量来近似地刻划它．反之若某个连续型随机变量在一个区间上连续取值（在此区间上密度大于 0），可以将此区间划分成若干子区间，在每个子区间上取一个值（譬如说子区间的中点）作为代表，用这个子区间上的概率表示这个代表值的概率．用这样一个离散分布去近似原来的连续分布．

3.1.3 随机向量

定义 3.1.10 设在同一样本空间 Ω 上定义的 n 个随机变量

$X_1,\cdots,X_n$ 构成一个向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$，称 $\boldsymbol{X}$ 为一个 **n 维随机向量**(n-dimensional random vector)，其中 X_i 称为 $\boldsymbol{X}$ 的第 i 个**分量**(component).

在一个试验中，人们有时关心的是多个数量指标以及其间的关系. 例如，从一群同年龄的儿童中抽出一个，测量他(她)的身高，体重和血色素这三个数量指标，便得到一个三维随机向量. 这里我们要特别强调，所说的随机向量中的各个分量必须是同一样本空间 Ω 上的随机变量.

n 维随机向量 $\boldsymbol{X}$ 的可能值是任一 n 元实数组 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$. 在上述例中假定 X_1＝身高(cm)，X_2＝体重(kg)，X_3＝血色素(g)，则 $\boldsymbol{X}=(115,20,14)$表示所抽测儿童的身高为 115cm，体重 20kg，血色素 14g.

定义 3.1.11 设 $\boldsymbol{X}$ 为 Ω 上的一个 n 维随机向量，对 n 维实空间 $\mathbf{R}^n$ 上的任一集合 A，定义

$$P(\boldsymbol{X} \in A) = P(\{\omega | \boldsymbol{X}(\omega) \in A\}),$$

称为 $\boldsymbol{X}$ 属于 A 的概率.

对 n 维随机向量，有时只考虑它的一个或几个分量的取值情况，这就相当于其它分量可以随意取值. 例如$\{X_1\in[0,1]\}=\{(x_1,\cdots,x_n),x_1\in[0,1],x_i\in\mathbf{R}^1,i=2,\cdots,n\}$. 这样的集合称为**柱形集**.

离散型 n 维随机向量的每个分量都是离散型随机变量.

定义 3.1.12 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$为离散型 n 维随机向量，记 X_i 的取值范围为$(x_{i_1},x_{i_2},\cdots)$，记

$$p_{j_1j_2\cdots j_n} = P(X_1 = x_{1j_1},X_2 = x_{2j_2},\cdots,X_n = x_{nj_n}),$$

则$\{(x_{1j_1},x_{2j_2},\cdots,x_{nj_n}),p_{j_1j_2\cdots j_n}\}$称为 $\boldsymbol{X}$ 的分布列.

二维离散型随机向量的分布列可用列表的方式表达. 设(X,Y)为二维随机向量，X 的取值范围为$(x_1,x_2,\cdots,x_n)$，Y 的取值范围为$(y_1,y_2,\cdots,y_m)$，记 $p_{ij}=P(X=x_i,Y=y_j)$，则可列表如表 3.1.

表 3.1

X \ Y	y_1	……	y_j	……	y_m	$\sum$
x_1	p_{11}	……	p_{1j}	……	p_{1m}	p_{1+}
⋮	⋮		⋮		⋮	
x_i	p_{i1}	……	p_{ij}	……	p_{im}	p_{i+}
⋮	⋮		⋮		⋮	
x_n	p_{n1}	……	p_{nj}	……	p_{nm}	p_{n+}
$\sum$	p_{+1}	……	p_{+j}	……	p_{+m}	1

$p_{i+}=\sum_{j=1}^{m}p_{ij},\quad p_{+j}=\sum_{i=1}^{n}p_{ij}.$

在表 3.1 中，中间列出 $p_{ij}=P(X=x_i,Y=y_j)$，下面一行和右边一列分别为每一列或每一行概率的求和，另有含义（见 3.2）.

定义 3.1.13 设 n 维随机向量 $\boldsymbol{X}$ 可在 $\mathbf{R}^n$ 中的一个区域内连续取值且存在 n 元实函数 $f(x_1,\cdots,x_n)$，对 $\mathbf{R}^n$ 中任一区域 A 有

$$P(\boldsymbol{X}\in A)=\int_A\cdots\int f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

则称 $\boldsymbol{X}$ 为连续型随机向量，称 $f(x_1,\cdots,x_n)$ 为 n 元**概率密度函数**，简称为 n 元**密度**.

推论 3.1.14 设 $f(x_1,\cdots,x_n)$ 为 n 元密度，则

(1) $f(x_1,\cdots,x_n)\geqslant 0$；

(2) $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1.$

混合型随机向量也会在实际中遇到，它是指随机向量的某些分量是离散型的，而另一些是连续型的，或某些分量本身是混合型的.

3.2 分布函数

3.2.1 一元分布函数

定义 3.2.1 设 X 为一随机变量，对于实函数 $F(x)$：

$$F(x) = P(X \leqslant x) = P(X \in (-\infty, x]).$$

称为随机变量 X 的**分布函数**(distribution function)，又称为**概率累积函数**(probability cumulative function).

分布函数有如下的性质.

性质 3.2.2 (1) $0 \leqslant F(x) \leqslant 1$；

(2) 单调性

$F(x) \leqslant F(y)$，当 $x < y$；

(3) 右连续性

$F(x-) = \lim\limits_{y \to x^+} F(y) = F(x)$；

(4) $F(-\infty) = \lim\limits_{y \to -\infty} F(y) = 0, F(+\infty) = \lim\limits_{y \to +\infty} F(y) = 1$；

(5) $P(x < X \leqslant y) = F(y) - F(x)$，当 $x < y$；

(6) $P(X = x) = F(x) - F(x+) = F(x) - \lim\limits_{y \to x^-} F(y)$.

分布函数的上述性质可以利用概率的性质加以验证. 由性质(5)和(6)可以知道，随机变量 X 在任一区间上的概率可由分布函数的值(或极限值)表示. 例如，

$$\left.\begin{aligned} P(a \leqslant X \leqslant b) &= F(b) - F(a+) \\ P(x < X) &= 1 - F(x) \end{aligned}\right\}. \tag{3.1}$$

由此可知，假定一个实数集 A 是由一些区间经过有限次或可列无穷多次的集合运算(并、交、差等)得到的，则 $P(X \in A)$ 可由分布函数 $F(x)$ 的值或极限值表示，因此分布函数唯一地确定了随机变量的概率分布.

设 X 为离散型随机变量，有分布列 $\{(x_k, p_k), k = 1, 2, \cdots\}$，则

它的分布函数可表为

$$F(x)=\sum_{x_k\leqslant x}p_k, \tag{3.2}$$

其图形为一个阶梯形函数,如图 3.3. 图 3.1 为离散形随机变量的分布函数.

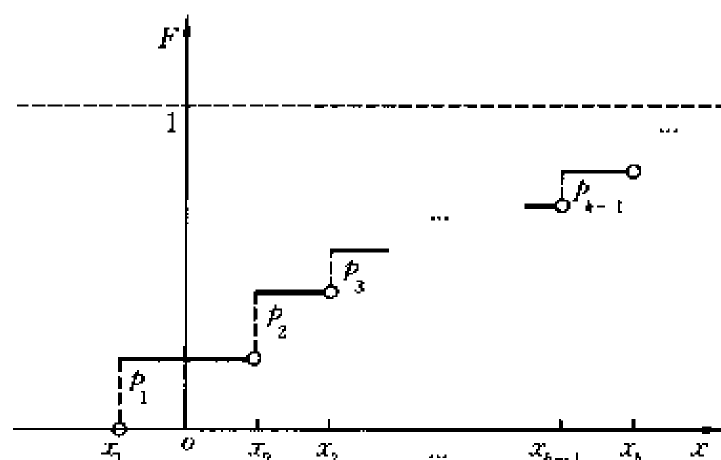


图 3.3

图中黑色水平线表示分布函数 $F(x)$.

设 X 为连续型随机变量,有密度 $f(x)$,则其分布函数 $F(x)$ 可表示为

$$F(x)=\int_{-\infty}^{x}f(u)\mathrm{d}u. \tag{3.3}$$

由积分的连续性可知如下推论.

推论 3.2.3 连续型随机变量 X 的分布函数 $F(x)$ 为连续函数;且对任意实数 x, $P(X=x)=0$.

分布函数 $F(x)$ 与密度 $f(x)$ 还有下述关系.

定理 3.2.4 设 $f(x)$ 为密度, $F(x)$ 为相应的分布函数,则

(1) $F(b)-F(a)=\int_a^b f(x)\mathrm{d}x$;

(2) 若 $f(x)$ 在 x_0 点连续,则

$$f(x_0)=F'(x)\big|_{x=x_0}.$$

3.2.2 多元分布函数

定义 3.2.5 设 $X=(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量，对任一 n 元实数组 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$，

$$\begin{aligned}F(x_1,x_2,\cdots,x_n) &= P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n)\\ &= P(\{X_1\in(-\infty,x_1]\}\cap\\ &\quad\cdots\cap\{X_n\in(-\infty,x_n]\}),\end{aligned}$$

则 $F(x_1,x_2,\cdots,x_n)$ 称为随机向量 X 的**联合分布函数**(joint distribution function)，又称为 n 元分布函数(n-dimensional distribution function).

性质 3.2.6 多元分布函数有下列的性质

(1) $0\leqslant F(x_1,x_2,\cdots,x_n)\leqslant 1$;

(2) 单调性

当 $x_i<y_i,i=1,\cdots,n$;

$$F(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n)\leqslant F(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots,x_n);$$

(3) 右连续性

$$\lim_{y_i\to x_i^+}F(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots,x_n)=F(x_1,\cdots,x_i,\cdots,x_n);$$

(4) $F(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)$

$$=\lim_{x_i\to-\infty}F(x_1,\cdots,x_i,\cdots,x_n)=0,$$

$$F(+\infty,\cdots,+\infty)=\lim_{\substack{x_i\to+\infty\\ i=1,\cdots,n}}F(x_1,\cdots,x_n)=1;$$

(5) $P(a_1<X_1\leqslant b_1,a_2<X_2\leqslant b_2,\cdots,a_n<X_n\leqslant b_n)$

$$\begin{aligned}&=F(b_1,b_2,\cdots,b_n)-\sum_{i=1}^{n}F(b_1,\cdots,b_{i-1},a_i,b_{i+1},\cdots,b_n)\\ &\quad+\sum_{i<j}F(b_1,\cdots,b_{i-1},a_i,b_{i+1},\cdots,b_{j-1},a_j,b_{j+1},\cdots,b_n)\\ &\quad-\cdots+(-1)^nF(a_1,a_2,\cdots,a_n).\end{aligned}$$

由性质(5)易见，n 维随机向量的概率分布由其(联合)分布函

数唯一确定.

离散型随机向量的分布函数可由其分布列来表示,例如设(X,Y)为二维向量,其分布列为$\{(x_i,y_j),p_{ij},i=1,2,\cdots,j=1,2,\cdots\}$,则

$$F(x,y)=\sum_{\substack{x_i<x\\ 且 y_j<y}} p_{ij}. \tag{3.4}$$

连续型随机向量的分布函数$F(x_1,\cdots,x_n)$可由其密度函数的积分得到:

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(u_1,\cdots,u_n)\mathrm{d}u_1\cdots\mathrm{d}u_n. \tag{3.5}$$

推论 3.2.7 设$f(x_1,\cdots,x_n)$为n元密度,$F(x_1,\cdots,x_n)$为相应的n元分布函数,则

(1)
$$P(a_i<X_i\leqslant b_i,i=1,\cdots,n)$$
$$=\int_{a_1}^{b_1}\cdots\int_{a_n}^{b_n} f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n;$$

(2) 在$f(x_1,\cdots,x_n)$的连续点上有

$$f(x_1,\cdots,x_n)=\frac{\partial^n F}{\partial x_1\cdots\partial x_n}.$$

3.2.3 边缘分布

定义 3.2.8 设$\boldsymbol{X}=(X_1,\cdots,X_n)$为$n$维随机向量,它的任意$k$个分量$X_{r_1},\cdots,X_{r_k}(1\leqslant k\leqslant n-1,1\leqslant r_1<r_2<\cdots<r_k\leqslant n)$构成一个$k$维随机向量$(X_{r_1},X_{r_2}\cdots,X_{r_k})$,其概率分布称为$\boldsymbol{X}$的一个边缘分布,其分布函数$F_{r_1\cdots r_k}(x_{r_1},\cdots,x_{r_k})$称为**边缘分布函数**(marginal distribution).

推论 3.2.9 设$F(x_1,\cdots,x_n)$为n元分布函数,则它的边缘分布函数为

$$F_{r_1r_2\cdots r_k}(x_{r_1},x_{r_2},\cdots,x_{r_k})=\lim_{\substack{x_j\to\infty\\ j\neq r_1,r_2,\cdots,r_k}} F(x_1,x_2,\cdots,x_n).$$

边缘分布函数 $F_{r_1r_2\cdots r_k}(x_{r_1},x_{r_2},\cdots,x_{r_k})$是一个 k 元分布函数，因而具有多元分布函数的性质(3.2.6). 一个 n 元分布函数可以产生$\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n-1}=2^n-2$ 个边缘分布函数.

对离散型的 n 维随机向量，其任意 k 个分量的边缘分布列可以由固定相应的下标而对其余的下标求和来得到. 例如对二元随机向量(X,Y)，设其分布列为$\{(x_i,y_j),p_{ij},i=1,2\cdots,j=1,2\cdots\}$，则 X 及 Y 的边缘分布列分别为

$$\left.\begin{aligned} P(X=x_i)&=\sum_j p_{ij}=p_{i+},i=1,2,\cdots,\\ P(Y=y_j)&=\sum_i p_{ij}=p_{+j},j=1,2,\cdots. \end{aligned}\right\} \tag{3.6}$$

由表 3.1 可见，$p_{i+},i=1,2,\cdots$和 $p_{+j},j=1,2,\cdots$分别为该表的右边一列和下边一行. 因此称为**边缘分布**.

对连续型 n 维随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$，记其密度为 $f(x_1,\cdots,x_n)$，则其任意 k 个分量$(X_{r_1},\cdots,X_{r_k})$的边缘密度 $f_{r_1r_2\cdots r_k}(x_{r_1},\cdots,x_{r_k})$为对 $f(x_1,\cdots,x_n)$的其余 $n-k$ 个变元从 $-\infty$到$+\infty$求积分得到. 例如设二维随机向量(X,Y)有联合密度 $f(x,y)$则它们各自的边缘密度为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y,\quad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x. \tag{3.7}$$

相应的边缘分布函数为

$$\left.\begin{aligned} F_X(x)&=\int_{-\infty}^{x}f_X(u)\mathrm{d}u=\int_{-\infty}^{x}\int_{-\infty}^{\infty}f(u,y)\mathrm{d}y\mathrm{d}u,\\ F_Y(y)&=\int_{-\infty}^{y}f_Y(v)\mathrm{d}v=\int_{-\infty}^{y}\int_{-\infty}^{\infty}f(x,v)\mathrm{d}x\mathrm{d}v. \end{aligned}\right\} \tag{3.8}$$

例 3.2.10 设二元随机向量(X,Y)有联合密度

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}}\qquad -\infty<x<\infty,-\infty<y<\infty,$$

则 X 与 Y 的边缘密度分别为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\ -\infty<x<\infty,$$

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}},\ -\infty<y<\infty.$$

推论 3.2.11 n 元分布函数唯一地确定它的每一个边缘分布函数,但反之不真.

例 3.2.12 下面两个表分别给出了两个二维离散随机向量的分布列.

X \ Y	0	1	p_{i+}
0	0.25	0.25	0.5
1	0.25	0.25	0.5
p_{+j}	0.5	0.5	1

X \ Y	0	1	p_{i+}
0	0.2	0.3	0.5
1	0.3	0.2	0.5
p_{+j}	0.5	0.5	1

由表中可以看出,这两个二维随机变量有相同的边缘分布,但它们的联合分布是不同的.

3.3 条件分布与随机变量的独立性

3.3.1 条件分布

当同时研究多个随机变量时,变量之间的相互影响,相互依赖的关系是一个重要的研究内容.可以从不同的角度来研究这个问题(例如在 4.4 中从相关和回归的角度来研究).本节研究条件分布的概念.设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量,假定已知其中一部分分量的值,在此条件下,其余分量的条件概率分布是什么?

首先设 X 为离散型随机变量,设对某个 $k<n$,已知 $X_1=x_{1j_1}$, $X_2=x_{2j_2},\cdots,X_k=x_{kj_k}$,则在此条件下,$(X_{k+1},X_{k+2},\cdots,X_n)$的条件

概率分布列为

$$P(X_{k+1}=x_{k+1\,j_{k+1}},\cdots,X_n=x_{nj_n}|X_1=x_{1j_1},\cdots,X_k=x_{kj_k})$$
$$=P(X_1=x_{1j_1},\cdots,X_n=x_{nj_n})/P(X_1=x_{1j_1},\cdots,X_k=x_{kj_k})$$
$$=p_{j_1\cdots j_n}/p_{j_1\cdots j_k+\cdots+} \qquad (3.9)$$

其中 $p_{j_1\cdots j_n}$如定义 3.1.12 所定义,$p_{j_1\cdots j_k+\cdots+}$ 为 $p_{j_1\cdots j_n}$对后 $n-k$ 个下标求和.

特别对二维离散型随机向量(X,Y),设其分布列为$\{(x_i,y_j),p_{ij},i=1,2,\cdots,j=1,2,\cdots\}$,则给定 $X=x_i$ 时 Y 的条件概率分布列为

$$P(Y=y_j|X=x_i)=p_{ij}/p_{i+} \quad j=1,2,\cdots, \qquad (3.10)$$

在表 3.1 中这个条件分布列可由中间第 i 行除以最右边一列的第 i 个值得到.

由条件分布列就可以得到条件分布函数:给定 $X=x_i$,Y 的条件分布函数为

$$F_{Y|X}(y|x_i)=\sum_{y_j\leqslant y}P(Y=y_j|X=x_i)=\sum_{y_j\leqslant y}p_{ij}/p_{i+}. \qquad (3.11)$$

但对连续型随机向量,直接由条件概率公式去定义条件分布是行不通的.设(X,Y)为二维连续型随机变量,有密度 $f(x,y)$,我们定义

$$P(Y\leqslant y|X=x)=\lim_{\Delta\to 0}\frac{P(x-\Delta<X\leqslant x+\Delta,Y\leqslant y)}{P(x-\Delta<X\leqslant x+\Delta)}$$
$$=\lim_{\Delta\to 0}\frac{\frac{1}{2\Delta}\int_{x-\Delta}^{x+\Delta}\int_{-\infty}^{y}f(x,y)\mathrm{d}x\mathrm{d}y}{\frac{1}{2\Delta}\int_{x-\Delta}^{x-\Delta}\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y\mathrm{d}x}. \qquad (3.12)$$

若 $f(x,y)$ 在点 (x,y) 处连续,则上式右端的极限值为 $\int_{-\infty}^{y}f(x,y)\mathrm{d}y\Big/\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y$.

定义 3.3.1 设二维连续型随机向量(X,Y)有密度 $f(x,y)$，则定义给定 $X=x$ 时 Y 的条件分布函数为

$$F_{Y|X}(y|x)=\int_{-\infty}^{y}\left[f(x,y)\Big/\int_{-\infty}^{\infty}f(x,v)\mathrm{d}v\right]\mathrm{d}y$$
$$=\int_{-\infty}^{y}f_{Y|X}(y|x)\mathrm{d}y,$$

其中

$$f_{Y|X}(y|x)=f(x,y)\Big/\int_{-\infty}^{\infty}f(x,v)\mathrm{d}v.$$

称为条件概率密度.

推论 3.3.2 在定义 3.3.1 的条件与记号下，有

(1) $f(x,y)=f_X(x)f_{Y|X}(y|x)$.

其中 $f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y$ 为 X 的边缘密度；

(2) Y 的边缘分布函数为

$$F_Y(y)=\int_{-\infty}^{\infty}F_{Y|X}(y|x)f_X(x)\mathrm{d}x=E[F_{Y|X}(y|x)].$$

例 3.3.3 设二维随机向量(X,Y)服从在单位圆上的均匀分布，即有密度

$$f(x,y)=\begin{cases}\dfrac{1}{\pi} & \text{当 } x^2+y^2<1,\\ 0, & \text{其它}.\end{cases}$$

则 X 的边缘密度为

$$f_X(x)=\frac{2}{\pi}\sqrt{1-x^2},\quad |x|<1.$$

故当给定 $X=x(|x|<1)$时，Y 的条件密度为

$$f_{Y|X}(y|x)=\frac{1}{2\sqrt{1-x^2}},\quad |y|<\sqrt{1-x^2}.$$

所以当给定 $X=x$，Y 的条件分布为区间$(-\sqrt{1-x^2},\sqrt{1-x^2})$上的均匀分布.

由定义 3.3.1，可得到当给定 $Y=y$ 时，X 的条件分布函数和条件密度分别为

$$\left.\begin{aligned} F_{X|Y}(x|y) &= \int_{-\infty}^{x}\left[f(x,y)\Big/\int_{-\infty}^{\infty}f(u,y)\mathrm{d}u\right]\mathrm{d}x, \\ f_{X|Y}(x|y) &= f(x,y)\Big/\int_{-\infty}^{\infty}f(u,y)\mathrm{d}u. \end{aligned}\right\} \tag{3.13}$$

将定义 3.3.1 推广到一般 n 维随机变量的场合有

定义 3.3.4 设 $X=(X_1,\cdots,X_n)$ 为 n 维连续型随机向量，有密度 $f(x_1,\cdots,x_n)$，则对 $k<n$，给定 $X_1=x_1,\cdots,X_k=x_k$ 时，$(X_{k+1},\cdots,X_n)$ 的条件密度为

$$\begin{aligned} & f_{X_{k+1}\cdots X_n|X_1\cdots X_k}(x_{k+1}\cdots x_n|x_1\cdots x_k) \\ = {} & f(x_1\cdots x_n)\Big/\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_k,v_{k+1},\cdots,v_n)\mathrm{d}v_{k+1}\cdots\mathrm{d}v_n, \end{aligned} \tag{3.14}$$

条件分布函数为

$$\begin{aligned} & F_{X_{k+1}\cdots X_n|X_1\cdots X_k}(x_{k+1},\cdots,x_n|x_1,\cdots,x_k) \\ = {} & \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{X_{k+1},\cdots,X_n|X_1\cdots X_k}(x_{k+1},\cdots,x_n|x_1,\cdots,x_k)\mathrm{d}x_{k+1},\cdots,\mathrm{d}x_n \end{aligned} \tag{3.15}$$

3.3.2 随机变量的独立性

定义 3.3.5 设 $X_1,X_2,\cdots,X_n$ 为一组随机变量，若对任意 n 个实数集合 $A_1,A_2,\cdots,A_n$ 有

$$\begin{aligned} & P(X_1\in A_1,X_2\in A_2,\cdots,X_n\in A_n) \\ & = P(X_1\in A_1)P(X_2\in A_2)\cdots P(X_n\in A_n), \end{aligned}$$

则称 $X_1,X_2,\cdots,X_n$ 为相互独立的随机变量.

随机变量间的独立性的一个重要来源是随机试验的独立性. 例如，设一个试验 E 是由两个独立的试验 E_1 和 E_2 复合而成(见 2.3.2)，设随机变量 X 和 Y 分别只与 E_1 和 E_2 的试验结果有关，

则 X 与 Y 是相互独立的随机变量.

定理 3.3.6 设 $X_1,X_2,\cdots,X_n$ 为一组随机变量,它们的联合分布函数为 $F(x_1,x_2,\cdots,x_n)$,各自的边缘函数分别为 $F_1(x_1)$,$F_2(x_2),\cdots,F_n(x_n)$,则 $X_1,X_2,\cdots,X_n$ 相互独立的充分必要条件为

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n).$$

定理 3.3.7 设 $X_1,X_2,\cdots,X_n$ 为一组离散型随机变量,则 $X_1,X_2,\cdots,X_n$ 相互独立的充分必要条件为对任一组可能值 x_{1i_1},$x_{2i_2},\cdots,x_{ni_n}$ 有

$$\begin{aligned}&P(X_1=x_{1i_1},X_2=x_{2i_2},\cdots,X_n=x_{ni_n})\\=&P(X_1=x_{1i_1})P(X_2=x_{2i_2}),\cdots,P(X_n=x_{ni_n}).\end{aligned}$$

定理 3.3.8 设 $X_1,X_2,\cdots,X_n$ 为一组连续型随机变量,它们的联合密度为 $f(x_1,x_2,\cdots,x_n)$,各自的边缘密度分别为 $f_1(x_1)$,$f_2(x_2),\cdots,f_n(x_n)$,则 $X_1,X_2,\cdots,X_n$ 相互独立的充分必要条件为

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)f_2(x_2)\cdots f_n(x_n).$$

推论 3.3.9 设 X,Y 为两个随机变量,则 X 与 Y 相互独立,当且仅当给定 X 时,Y 的条件分布与 Y 的(边缘)分布相同;当且仅当给定 Y 时,X 的条件分布与 X 的(边缘)分布相同.

例 3.3.10 在一个"成功"概率为 p 的伯努利试验中,定义

$$X_i=\begin{cases}1, & \text{当第 } i \text{ 次试验成功},\\ 0, & \text{当第 } i \text{ 次试验失败},\end{cases}\quad i=1,\cdots,n,$$

则 $X_1,\cdots,X_n$ 为相互独立的随机变量,且每个 X_i 有分布 $P(X_i=1)=p,P(X_i=0)=1-p,i=1,\cdots,n.$

例 3.3.11 在例 3.2.10 中易见 $f(x,y)=f_X(x)f_Y(y)$,故 X 与 Y 为相互独立的随机变量.

以上讨论的是随机变量间的独立性,进一步可定义随机向量间的独立性.为表达简便计,只考虑两个随机向量.

定义 3.3.12 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 与 $\boldsymbol{Y}=(Y_1,\cdots,Y_m)$ 分别为 n

维和 m 维随机向量. 若对任一 $\mathbf{R}^n$ 中的集合 A 和 $\mathbf{R}^m$ 中的集合 B 有

$$P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B) = P(\boldsymbol{X} \in A) \cdot P(\boldsymbol{Y} \in B),$$

则称随机向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立.

注 (1) 在上述定义中不涉及任一随机向量内部各分量之间的独立性问题.

(2) 上述定义仿照定义 3.2.6 可推广到多个随机向量相互独立的场合.

定理 3.3.13 设 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 分别为 n 维和 m 维随机向量, 则 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立, 当且仅当

$$F(x_1, \cdots, x_n; y_1, \cdots, y_m) = F_X(x_1, \cdots, x_n) \cdot F_Y(y_1, \cdots, y_m).$$

其中 F 为 $(\boldsymbol{X}, \boldsymbol{Y})$ 的联合分布函数, F_X 与 F_Y 分别为 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的(边缘)分布函数.

3.4 随机变量的函数及其分布

随机变量(或向量)的函数仍然是随机变量. 设 X 为随机变量(向量), $g(x)$ 为实函数, 则 $Y=g(X)$ 为一随机变量.

对任一实数集 A, 记 $g^{-1}(A)=\{x \mid g(x) \in A\}$,

$$P(Y \in A) = P(g(X) \in A) = P\{X \in g^{-1}(A)\}. \quad (3.16)$$

公式(3.16)是计算随机变量函数分布的基本公式.

设 X 为离散型随机变量, 取值于 $\{x_1, x_2, \cdots\}$, 则对 $Y=g(X)$ 有

$$P(Y \in A) = \sum_{g(x_i) \in A} P(X = x_i). \quad (3.17)$$

此公式可推广到 X 为离散型随机向量的场合.

定理 3.4.1 设 X_1 和 X_2 为相互独立的取值于非负整数集合的随机变量, $Y=X_1+X_2$, 则

$$P(Y=k)=\sum_{i=0}^{k}P(X_1=i)\cdot P(X_2=k-i),k=0,1,2,\cdots.$$

此公式称为**离散卷积**(discrete convolution)公式.

设 X 为连续型随机变量,有密度函数 $f(x)$,则对 $Y=g(X)$,Y 的分布函数可表为

$$F_Y(y)=P(Y\leqslant y)=\int_{g(x)\leqslant y}f(x)\mathrm{d}x. \qquad (3.18)$$

Y 的密度可表示为

$$f_Y(y)=F'_Y(y). \qquad (3.19)$$

上述公式可推广到 X 为 n 维连续型随机变量的场合.

定理 3.4.2 设 X 为连续型随机变量,其密度为 $f(x)$,又设 $g(x)$ 为严格单调函数,且反函数可导,则 $Y=g(x)$ 有密度

$$f_Y(y)=f(g^{-1}(y))\,|(g^{-1}(y))'|.$$

定理 3.4.3 设 X_1 和 X_2 为独立连续型随机变量,分别有密度 $f_1(x_1)$ 和 $f_2(x_2)$,$Y=X_1+X_2$,则 Y 有密度

$$\begin{aligned}f_Y(y)&=\int_{-\infty}^{\infty}f_1(y-v)f_2(v)\mathrm{d}v\\&=\int_{-\infty}^{\infty}f_1(u)f_2(y-u)\mathrm{d}u.\end{aligned}$$

此公式可记作 $f_Y=f_1*f_2=f_2*f_1$,称为**积分卷积**(integral convolution)公式.

下面考虑多个随机变量函数的联合分布.

定理 3.4.4 设 $X_1,X_2,\cdots,X_n$ 为连续型随机变量,有联合密度 $f(x_1,x_2,\cdots,x_n)$,设 $Y_i=g_i(X_1,\cdots,X_n)$,$i=1,\cdots,n$. 若变换 $(x_1,x_2,\cdots,x_n)\to(g_1(x_1,\cdots,x_n),\cdots,g_n(x_1,\cdots,x_n))$ 有反变换存在,记作 $x_i=\varphi_i(g_1,\cdots,g_n)$,$i=1,\cdots,n$,其雅可比行列式

$$J=\begin{vmatrix}\frac{\partial \varphi_1}{\partial g_1} & \cdots & \frac{\partial \varphi_1}{\partial g_n}\\ \cdots & \cdots & \cdots \\ \frac{\partial \varphi_n}{\partial g_1} & \cdots & \frac{\partial \varphi_n}{\partial g_n}\end{vmatrix}=\begin{vmatrix}\frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n}\\ \cdots & \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n}\end{vmatrix}^{-1}.$$

处处存在，则$(Y_1,\cdots,Y_n)$有联合密度

$$f_Y(y_1,\cdots,y_n)=f(\varphi_1(y_1,\cdots,y_n),\cdots,\varphi_n(y_1,\cdots,y_n))\cdot|J|.$$

注 在上述定理中反变换的存在，即$(x_1,\cdots,x_n)$与$(g_1,\cdots,g_n)$的1-1对应是一个关键的条件.此条件可进一步推广.设在$\mathbf{R}^n$的m个互不相交的区域$\Sigma_1,\Sigma_2,\cdots,\Sigma_m$上，对$(x_1,\cdots,x_n)\in\Sigma_j$，$(x_1,\cdots,x_n)$与$(g_1,\cdots,g_n)$的1-1对应存在，记$x_i=\varphi_i^{(j)}(g_1,\cdots,g_n)$，$i=1,\cdots,n$，且相应的Jacobi行列式$J_j$在$\Sigma_j$上处处存在，$j=1,2,\cdots,m$，则$(Y_1,\cdots,Y_n)$有联合密度

$$f_Y(y_1,\cdots,y_n)=\sum_{j=1}^{m}f(\varphi_1^{(j)}(y_1,\cdots,y_n),\cdots,\varphi_n^{(j)}(y_1,\cdots,y_n))\cdot|J_j|.\quad(3.20)$$

例 3.4.5 设$X_1,\cdots,X_n$相互独立，有相同的密度

$$f(x)=\mathrm{e}^{-x},\ 0<x<\infty,$$

设$Y_k=\sum_{i=1}^{k}X_i,k=1,\cdots,n$. 则定理3.4.4的条件成立，雅可比行列式$|J|=1$，于是$(Y_1,\cdots,Y_n)$有密度

$$f_Y(y_1,\cdots,y_n)=\mathrm{e}^{-y_n},\ 0<y_1<y_2<\cdots<y_n<\infty.$$

例 3.4.6 设X与Y相互独立且均服从标准正态分布$N(0,1)$（见定义6.4.5），设$U=X^2+Y^2$，$V=X/Y$，则U与Y相互独立.

证 (X,Y)的联合密度为

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}},$$

$$-\infty<x<+\infty,\ -\infty<y<+\infty.$$

又分别在(X,Y)的两个取值区域 $\Sigma_1=\mathbf{R}^1\times(0,\infty)$，$\Sigma_2=\mathbf{R}^1\times(-\infty,0)$上，$(X,Y)$与$(U,V)$是 1-1 对应的，相应的雅可比行列式为

$$J_1=J_2=-\frac{1}{2(1+V^2)}.$$

故(U,V)的联合密度为

$$h(u,v)=2\left[\frac{1}{2\pi}\mathrm{e}^{-\frac{u^2}{2}}\frac{1}{2(1+v^2)}\right]=\frac{1}{2\pi}\mathrm{e}^{-\frac{u}{2}}\Big/(1+v^2),$$

$$u>0,\ -\infty<v<\infty.$$

由 $h(u,v)$可算出 U 和 V 的边缘密度分别为

$$f_U(u)=\frac{1}{2}\mathrm{e}^{-\frac{u}{2}}\qquad 0<u,$$

$$f_V(v)=\frac{1}{\pi(1+v^2)}\quad -\infty<v<\infty.$$

故 $h(u,v)=f_U(u)f_V(v)$，由定理 3.3.8 知 U 与 V 相互独立.

定义 3.4.7 设随机变量 $X_1,X_2,\cdots,X_n$ 独立同分布，定义由 $X_1,\cdots,X_n$ 生成的**顺序统计量**(order statistics)为 $X_1,\cdots,X_n$ 的一个重排 $X_{(1)},\cdots,X_{(n)}$满足

$$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}.$$

定理 3.4.8 设 $X_1,\cdots,X_n$ 为独立同分布随机变量，有公共分布函数 $F(x)$和密度 $f(x)$，$X_{(1)},\cdots,X_{(n)}$为由它们产生的顺序统计量，则

(1) $X_{(1)},\cdots,X_{(n)}$的联合密度为

$$f(x_1,x_2,\cdots,x_n)=\begin{cases}n!f(x_1,\cdots,x_n) & x_1\leqslant x_2\leqslant\cdots\leqslant x_n,\\ 0 & \text{其它}.\end{cases}$$

(2) 对 $1\leqslant r_1<r_2<\cdots<r_k\leqslant n$，$(X_{(r_1)},X_{(r_2)},\cdots,X_{(r_k)})$的密度为

$$f_{r_1\cdots r_k}(x_1,\cdots,x_k)$$
$$=n!\prod_{i=1}^{k}f(x_i)\prod_{i=0}^{k}\left\{\frac{[F(x_{i+1})-F(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!}\right\},$$

$x_1 \leqslant x_2 \leqslant \cdots \leqslant x_k$. 其中 $x_0 = -\infty, x_{k+1} = +\infty, r_0 = 0, r_{k+1} = n+1$.

(3) $X_{(r)}$ 的密度为 $f_r(x) = \frac{n!}{(r-1)!\,(n-r)!} F^{r-1}(x)[1-F(x)]^{n-r} f(x)$. 特别 $X_{(1)}$ 的密度为 $f_1(x) = n[1-F(x)]^{n-1} f(x)$, $X_{(n)}$ 的密度为 $f_n(x) = nF^{n-1}(x) f(x)$.

推论 3.4.9 设 $X_1, \cdots, X_n$ 独立有相同连续分布函数 $F(x)$, $X_{(1)}, \cdots, X_{(n)}$ 为顺序统计量, 记 $Y_1 = F(X_{(1)}), Y_i = F(X_{(i)}) - F(X_{(i-1)}), i = 2, \cdots, n, Y_{n+1} = 1 - F(X_{(n)})$, 则 $(Y_1, \cdots, Y_n, Y_{n+1}) \sim D_n(1, \cdots, 1)$ (见定义 6.6.8).

4 随机变量的数字特征

分布函数唯一地决定了随机变量(向量)的全部数字特征. 而随机变量的某些重要特征又可用少量的数值来刻画. 这样的数值统称为随机变量或分布函数的**数字特征**.

4.1 数学期望,位置参数

许多数字特征与分布函数的斯蒂尔切斯(Stieltjes)**积分**有关. 定义如下

定义 4.1.1 设随机变量 X 有分布函数 $F(x)$,对任一实函数 $h(x)$,定义它关于 $F(x)$的斯蒂尔切斯积分 $\int_{-\infty}^{\infty} h(x)\mathrm{d}F(x)$ 如下:

(1) 若 X 为离散型,有分布列 $\{(x_i, p_i), i=1,2,\cdots\}$, 且 $\sum_i |h(x_i)| p_i < \infty$, 则

$$\int_{-\infty}^{\infty} h(x)\mathrm{d}F(x) = \sum_i |h(x_i)| p_i;$$

(2) 若 X 为连续型,有密度 $f(x)$, 且 $\int_{-\infty}^{\infty} |h(x)| f(x)\mathrm{d}x < \infty$, 则

$$\int_{-\infty}^{\infty} h(x)\mathrm{d}F(x) = \int_{-\infty}^{\infty} h(x)f(x)\mathrm{d}x.$$

注 当 $\sum_i |h(x_i)| p_i < \infty$ 或 $\int_{-\infty}^{\infty} |h(x)| f(x)\mathrm{d}x < \infty$ 的条件不成立时,相应的斯蒂尔切斯积分无定义.

4.1.1 数学期望

定义 4.1.2 设随机变量 X 有分布函数 $F(x)$，定义其**数学期望**(mathematical expectation)为

$$E(X)=\int_{-\infty}^{\infty}x\mathrm{d}F(x).$$

当上式右端积分无定义时，数学期望亦无定义. $E(X)$又称为**均值**(mean).

数学期望是一个刻画随机变量的取值"中心"的量. 例如，假定 X 的分布关于 μ 对称，即对任意的 x，$P(X-\mu>x)=P(X-\mu<-x)$，则 $E(X)=\mu$.

从频率逼近概率的角度来看，大数定律(见第 9 章)证明了，对一串独立同分布的随机变量 $X_1,X_2,\cdots,X_n$，记它们共同的均值为 μ，则当 n 充分大时，$\frac{1}{n}\sum_1^n X_i$ 逼近 μ. 这个事实解释了"数学期望"这个词的含义.

例 4.1.3 设 $X\sim B(n,p)$，二项分布，即 $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$，$k=0,1,\cdots,n$，则

$$E(X)=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}=p.$$

例 4.1.4 设随机变量 X 服从柯西(Cauchy)分布，即有密度 $f(x)=\frac{1}{\pi(1+x^2)}$ $-\infty<x<\infty$，则 $E(X)$不存在，因为

$$\int_{-\infty}^{\infty}\frac{|x|}{\pi(1+x^2)}\mathrm{d}x=\infty.$$

定义 4.1.5 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$为随机向量，且 $E(X_i)$，$i=1,2,\cdots,n$都存在，则定义 $\boldsymbol{X}$ 的数学期望(向量)为

$$E(\boldsymbol{X})=(E(X_1),E(X_2),\cdots,E(X_n)).$$

定理 4.1.6 设 X 有分布函数 $F(x)$，$Y=h(X)$，则

$$E(Y)=E(h(X))=\int_{-\infty}^{\infty}h(x)\mathrm{d}F(x),$$

当等式右端的积分存在.

此定理说明，当求 $E(Y)$时，不必先求出 Y 的分布函数，再由定义 4.1.2 进行计算，而可以通过直接求 $h(x)$关于 $F(x)$的斯蒂尔切斯积分得到.

例 4.1.7 设 $X\sim N(0,1)$，即有密度 $f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$，$-\infty<x<+\infty$，则

$$E(X^2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=1.$$

为定义多个随机变量函数的数学期望，下面引入多元斯蒂尔切斯积分的概念.

定义 4.1.8 设 n 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$有联合分布函数 $F(x_1,x_2,\cdots,x_n)$，对任一 n 元实函数 $h(x_1,\cdots,x_n)$，定义其斯蒂尔切斯积分 $\int\cdots\int h(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)$ 如下：

(1) 若 $\boldsymbol{X}$ 为离散型随机向量，且 $\sum|h(x_{1i_1},\cdots,x_{ni_n})|P_{i_1\cdots i_n}<\infty$，其中，$P_{i_1\cdots i_n}=P(X_1=x_{1i_1},\cdots,X_n=x_{ni_n})$，$\sum$ 为对$(X_1,\cdots,X_n)$的所有可能值求和，则

$$\int\cdots\int h(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)=\sum h(x_{1i_1},\cdots,x_{ni_n})P_{i_1\cdots i_n}.$$

(2) 若 $\boldsymbol{X}$ 为连续型随机向量，有密度 $f(x_1,\cdots,x_n)$，且 $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|h(x_1,\cdots,x_n)|f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n$ 存在，则

$$\int\cdots\int h(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)$$
$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h(x_1,\cdots,x_n)f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

定理 4.1.9 设 $X_1,\cdots,X_n$ 为随机变量，有联合分布函数

$F(x_1,\cdots,x_n)$，且 $Y=h(X_1,\cdots,X_n)$，则

$$E(Y)=\int\cdots\int h(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n).$$

例 4.1.10 设(X_1,X_2)有密度 $f(x,y)=\frac{1}{\pi}$，当 $x^2+y^2<1$，则 $E(X_1X_2)=\frac{1}{\pi}\iint\limits_{x^2+y^2<1}xy\mathrm{d}x\mathrm{d}y=0.$

性质 4.1.11 数学期望的性质

(1) 设 C 是常数，则 $E(C)=C$；

(2) 设 X 为随机变量，a,b 为常数，则

$$E(aX+b)=aE(X)+b;$$

(3) $E(X+Y)=E(X)+E(Y)$；

(4) 设 $X_1,X_2,\cdots,X_n$ 相互独立，则

$$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n).$$

例 4.1.12 设 $X_1,\cdots,X_n$ 独立同分布，$P(X_1=1)=p$，$P(X_1=0)=1-p$，则 $X=\sum_{i=1}^{n}X_i\sim B(n,p)$，故

$$E(X)=\sum_{i=1}^{n}E(X_i)=np.$$

4.1.2 条件期望

定义 4.1.13 设 X,Y 为两个随机变量，给定 Y 时，X 的条件分布函数为 $F(x|Y)$. 定义给定 Y 时 X 的**条件期望**(conditional expectation)为

$$E(X|Y)=\int_{-\infty}^{\infty}x\mathrm{d}F(x|Y).$$

条件期望又称为**条件均值**.

定理 4.1.14 设 X,Y 为随机变量，$g(x)$为一实函数，则

(1) $E[g(X)]=E[E(g(X)|Y)]$；

(2) $E(XY)=E[YE(X|Y)]$.

注 在上述定理中,注意到 $E(g(X)|Y)$ 为 Y 的函数,因而也是随机变量,也可以求期望.公式(1)的含义是:求 $g(X)$ 的期望可以分两步走,先对某个已知变量(Y)求 $g(X)$ 的条件期望,最后再求条件期望的期望.

例 4.1.15 设某种昆虫的一条雌成虫产 k 个卵的概率为 $p_k=\frac{\lambda k}{k!}e^{-\lambda}, k=0,1,2,\cdots$,设每个卵孵化为幼虫的概率为 p,且卵的孵化相互之间是独立的.求平均每条雌成虫有多少幼虫.

解 设 X 为幼虫数,Y 为卵数,则给定 $Y=k$ 时,$X\sim B(k,p)$,故 $E(X|Y=k)=kp$(见例 4.1.12),所以 $E(X)=E(pY)=pE(Y)=p\lambda$.

4.1.3 位置参数

定义 4.1.16 设随机变量 X 的分布为 F,若由 F 决定的一个量 $\theta(X)$ 满足

$$\theta(X+C)=\theta(X)+C, \text{对任意常数} C,$$

则称 $\theta(X)$ 为 X 的(或 F 的)一个**位置参数**(location parameter).

位置参数在下述意义下刻划了分布的"位置":当随机变量平移了一个固定的量时,由分布决定的该参数也平移了相应的量.

由数学期望的性质 $E(X+C)=E(X)+C$ 知

推论 4.1.17 数学期望为一个位置参数.

除了数学期望之外,还有一些刻画随机变量分布"中心"的位置参数.

定义 4.1.18 对随机变量 X 及其分布函数 $F(x)$,定义**中位数**(median)为一实数 $m(X)$,满足

$$F(m(X))=P(X\leqslant m(X))\geqslant\frac{1}{2},$$

且 $$1-F(m(X))=P(x\geqslant m(X))\geqslant\frac{1}{2}.$$

性质 4.1.19 中位数的性质

(1) 中位数 $m(X)$一定存在，但不必唯一；

(2) 中位数 $m(X)$，唯一的充要条件为 $F(x)$在 $m(X)$点严格增；

(3) 设 $F(x)$连续，则

$$P(X\leqslant m(X))=\frac{1}{2}=P(X\geqslant m(X));$$

(4) 中位数 $m(X)$为位置参数，即

$$m(X+C)=m(X)+C,\quad 对任意常数 C.$$

定义 4.1.20 设随机变量 X 有密度 $f(x)$，且存在 x_0 满足

$$f(x_0)=\max_x f(x),$$

则称 x_0 为 X 的**众数**(mode).

推论 4.1.21 众数为位置参数.

定义 4.1.22 设 $F(x)$为分布函数，若对给定的 $p\in(0,1)$，实数 x_p 和 z_p 分别满足

$$x_p=\sup\{x:F(x)\leqslant p\},$$
$$z_p=\inf\{z:1-F(z)\leqslant p\}.$$

则称 x_p 为 F 的**下 α 分位数**(lower α-quantile)，而称 z_p 为 F 的**上 α 分位数**(upper α-quantile). 二者都可简称为**分位数**(quantile).

推论 4.1.23 分位数也是位置参数.

4.2 方差、刻度参数

4.2.1 方差

定义 4.2.1 设随机变量 X 有分布函数 $F(x)$，且

$\int_{-\infty}^{\infty} x^2 \mathrm{d}F(x) < \infty$，定义 X 的**方差**(variance)为

$$\sigma^2(X) = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 \mathrm{d}F(x).$$

定义**标准差**(standard deviation)为

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

方差(标准差)是用于刻画随机变量围绕均值的散布程度的量. 以正态分布 $N(\mu, \sigma^2)$ 为例，它的均值为 μ，方差为 σ^2，99.97% 的概率集中在区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 上. 因此，σ 越小，随机变量的取值就越集中在均值附近；反之，它向均值左右两边散开的程度就大.

性质 4.2.2　方差的性质

(1) $\sigma^2(X+c) = \sigma^2(X)$，对任意常数 c；

(2) $\sigma^2(aX) = a^2\sigma^2(X)$，对任意常数 a；

(3) $\sigma^2(X) = 0$，当且仅当存在常数 μ，使得

$$P(X = \mu) = 1;$$

(4) $\sigma^2(X) = E(X^2) - (EX)^2$；

(5) $\sigma^2(X) = \min\limits_{c \in \mathbf{R}} E(X - c)^2$.

(6) 设 $X_1, \cdots, X_n$ 为相互独立的随机变量，则

$$\sigma^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma^2(X_i).$$

例 4.2.3　在 n 重伯努利试验中，记

$$X_i = \begin{cases} 1, \text{当第 } i \text{ 次试验成功}, \\ 0, \text{当第 } i \text{ 次试验失败}. \end{cases} \quad i = 1, \cdots, n$$

则 $X_1, \cdots, X_n$ 相互独立，$P(X_i = 1) = p$，$P(X_i = 0) = 1 - p$. 故 $\sigma^2(X_i) = p(1-p)$. 令 $X = \sum_{1}^{n} X_i$，则 $\sigma^2(X) = np(1-p)$.

定义 4.2.4　设 X, Y 为二随机变量，定义给定 Y 时 X 的条件

方差为

$$\sigma^2(X|Y) = E\{[X - E(X|Y)]^2|Y\}.$$

即在给定 Y 时 X 的条件分布下求方差.

推论 4.2.5 $\sigma^2(X)=E[\sigma^2(X|Y)]+\sigma^2[E(X|Y)]$.

上述推论说明,当有另一个变量 Y 在对 X 起作用时,X 的方差可分为两部分:一部分是 Y 对 X 的系统影响部分($E(X|Y)$)的方差,另一部分是条件方差 $\sigma^2(X|Y)$的平均.

4.2.2 刻度参数

定义 4.2.6 设由随机变量 X 的分布所决定的参数 $\theta(X)$ 满足

$$\theta(aX) = a\theta(X),\text{对任意 } a > 0,$$

则称 $\theta(X)$为**刻度参数**(scale parameter).

一个合理的刻画随机变量散布度的数字特征应该是一个刻度参数. 这样当随机变量的值依单位选取不同而放大缩小若干倍时,该数字特征也以相同的比例变化.

推论 4.2.7 标准差 $\sigma(X)$是刻度参数.

定义 4.2.8 设 $m(X)$为 X 的中位数,则定义其**绝对差**(absolute deviation)为 $E|X-m(X)|$.

推论 4.2.9

(1) $E|X-m(X)|=\min\limits_{c} E|X-c|$.

(2) 绝对差 $E|X-m(X)|$为刻度参数.

定义 4.2.10 设对给定的 $p\in(0,1)$,记 $\alpha=1-p$ 随机变量 X 的上 $\frac{\alpha}{2}$ 分位数为 $z_{\frac{\alpha}{2}}$,下 $\frac{\alpha}{2}$ 分位数为 $x_{\frac{\alpha}{2}}$,定义其 **p 分位差**(quantile difference)为 $z_{\frac{\alpha}{2}}-x_{\frac{\alpha}{2}}$.

推论 4.2.11 p 分位差为刻度参数.

4.3 矩

定义 4.3.1 设随机变量 X 有分布函数 $F(x)$，

(1) 对给定的实数 $\gamma>0$，其 γ **阶绝对矩**(absolute moment of order γ)为

$$E|X|^{\gamma}=\int_{-\infty}^{\infty}|x|^{\gamma}\mathrm{d}F(x).$$

(2) 对给定的正整数 k，若 $E|X|^{k}$ 存在，其 k 阶**原点矩**(moment about origin)为

$$\alpha_k=E(X^k)=\int_{-\infty}^{\infty}x^k\mathrm{d}F(x).$$

(3) 对大于 1 的正整数 m，若 $E|X|^{m}$ 存在，其 m 阶**中心矩**(moment about centre)为

$$\mu_m=E(X-EX)^m=\int_{-\infty}^{\infty}(x-EX)^m\mathrm{d}F(x).$$

矩是广泛应用的一类数字特征，熟知的均值和方差分别为一阶原点矩和二阶中心矩.

定理 4.3.2 若 $0<q<p$，则$[E|X|^q]^{\frac{1}{q}}\leqslant[E|X|^p]^{\frac{1}{p}}$.

推论 4.3.3 中心矩或原点矩存在为有限的充分必要条件是同阶的绝对矩为有限.

由矩又可派生出一些有用的数字特征.

定义 4.3.4 设分布函数 $F(x)$有中心矩 $\mu_2=E(X-EX)^2$，$\mu_3=E(X-EX)^3$，其**偏度系数**(coefficient of skewness)为

$$C_s=\mu_3/\mu_2^{\frac{3}{2}}.$$

偏度系数是一个无量纲的量，它刻画分布函数的对称性. 当 $C_s>0$ 时，$F(x)$所表示的概率分布偏向均值的右侧，反之则偏向左侧.

定义 4.3.5 设分布函数有二阶中心矩 $\mu_2=E(X-EX)^2$ 和四阶中心矩 $\mu_4=E(X-EX)^4$，其**峰度系数**(kurtosis)为

$$C_k = \mu_4/\mu_2^2.$$

峰度 C_k 是一个无量纲的量，它用来刻画不同类型的分布的集中和分散的程度，但要注意这种比较是在使用相同刻度的条件下进行的.

定义 4.3.6 设 X 为随机变量，有分布函数 $F(x)$，则对任意实数 $a>0$ 及 b，称 $aX+b$ 与 X 有**同型分布**，其分布函数为 $F\left(\frac{x-b}{a}\right)$.

推论 4.3.7 同型分布有相同的峰度.

对不同类型的分布，可将随机变量标准化后再比较其密度函数或分布列的峰度.

定义 4.3.8 设随机变量 X 有均值 μ 和方差 σ^2，**其标准化随机变量**为

$$X^* = (X-\mu)/\sigma.$$

假定随机变量 X 是标准化的，其密度(或分布列)有一个单峰，则 C_k 愈小，其密度的单峰愈"陡峭"；C_k 愈大，其密度的单峰愈"平缓". 对正态分布来说 $C_k=3$. 因此，一个对称分布，其峰度系数愈接近 3，就愈接近正态分布.

4.4 相关与回归

相关(correlation)与**回归**(regression)是研究随机变量之间相互关系的两个重要概念，且彼此之间有紧密联系.

设 X 与 Y 为两个随机变量，并记

$$\mu_1 = E(X),\ \sigma_1^2 = \sigma^2(X);\ \mu_2 = E(Y),\ \sigma_2^2 = \sigma^2(Y).$$

首先考虑 Y 对 X 的影响. 所谓回归，就是用 Y 的一个函数

$g(Y)$去近似 X.

定理 4.4.1 在均方误差最小的意义下,$E(X|Y)$是 X 的最佳近似,即

$$E[X - E(X|Y)]^2 = \min_g E[X - g(Y)]^2.$$

定义 4.4.2 $E(X|Y)$称为 X 关于 Y 的**回归函数**(regression function).

对不同的分布,回归函数的形式是不同的,为便于计算,我们常用 Y 的一个线性函数去近似 X.

定义 4.4.3 设 $\beta_1 Y+\beta_0$ 满足

$$E[X - (\beta_1 Y + \beta_0)]^2 = \min_{a,b} E[X - (aY + b)]^2,$$

则称 $\beta_1 Y+\beta_0$ 为 X 的**线性回归函数**.

对二元正态分布,线性回归函数就是回归函数.

定理 4.4.4 设 X 关于 Y 的线性回归函数为 $\beta_1 Y+\beta_0$,则

$$\beta_1 = \rho\frac{\sigma_1}{\sigma_2}, \quad \beta_0 = \mu_1 - \beta_1\mu_2.$$

其中

$$\rho = E[(X - \mu_1)(Y - \mu_2)]/\sigma_1\sigma_2 = \sigma_{12}/\sigma_1\sigma_2,$$

且

$$E[X - (\beta_1 Y + \beta_0)]^2 = \sigma_1^2(1 - \rho^2).$$

定义 4.4.5 对随机变量 X,Y,定义 X 与 Y 的**协方差**(covariance)为

$$\begin{aligned}\text{cov}(X,Y) &= E[(X - EX)(Y - EY)] \\ &= E[(X - \mu_1)(Y - \mu_2)].\end{aligned}$$

X 与 Y 的**相关系数**(correlation coefficient)为

$$\rho(X,Y) = \text{cov}(X,Y)/\sqrt{\sigma^2(X)\sigma^2(Y)}.$$

以下简记 $\text{cov}(X,Y)=\sigma_{12}$,$\rho(X,Y)=\rho$.

协方差 σ_{12} 和相关系数 ρ 是刻画随机变量之间**相依性**

(interdependence)的数字特征. 由定义可见,σ_{12}与 ρ 有相同的符号. 当 $\sigma_{12}>0(\rho>0)$时,X 与 Y 有相同的变化趋势;当 $\sigma_{12}<0(\rho<0)$时,X 与 Y 有相反的变化趋势.

定义 4.4.6 当 $\sigma_{12}=\rho=0$ 时,称 X 与 Y **不相关**.

协方差与相关系数有以下性质:

性质 4.4.7 (1) $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$,

$$\rho(X,Y)=\rho(Y,X).$$

(2) $\mathrm{cov}(aX+b,Y)=a\mathrm{cov}(X,Y)$,

$\rho(aX+b,Y)=\rho(X,Y)$.

(3) $\mathrm{cov}(X_1+X_2,Y)=\mathrm{cov}(X_1,Y)+\mathrm{cov}(X_2,Y)$.

(4) $\mathrm{cov}(X,Y)=E(XY)-(EX)(EY)$.

定理 4.4.8 若 X 与 Y 独立,则 $\mathrm{cov}(X,Y)=\rho(X,Y)=0$.

由此定理可知,由随机变量的独立性可推出不相关性,但反过来结论不一定成立.

例 4.4.9 设 X,Y 的联合密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2<1,\\ 0, & \text{其它}.\end{cases}$$

则 $\mathrm{cov}(X,Y)=\rho(X,Y)=0$,因此 X,Y 不相关,但由例 3.3.3 知,X 与 Y 并不相互独立.

定理 4.4.10 柯西不等式

(1) $[\mathrm{cov}(X,Y)]^2\leqslant\sigma^2(X)\cdot\sigma^2(Y)$(或等价地 $\rho^2(X,Y)\leqslant1$).

(2) $[\mathrm{cov}(X,Y)]^2=\sigma^2(X)\cdot\sigma^2(Y)$(等价地,$\rho^2(X,Y)=1$).

当且仅当 X 与 Y 之间有严格线性关系.

相关系数是用来刻画随机变量间的线性相关程度的,特别对正态分布或接近正态分布的随机变量来说更是如此.

4.5 方差矩阵与相关矩阵

定义 4.5.1 设 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$，$\boldsymbol{Y}=(Y_1,\cdots,Y_m)^{\mathrm{T}}$ 为两个随机向量，$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的**协方差阵**(covariance matrix)为

$$\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y}) = (\sigma_{ij})_{n\times m},$$

其中 $\sigma_{ij}=\operatorname{cov}(X_i,Y_j)$，$i=1,\cdots,n$，$j=1,\cdots,m$.

性质 4.5.2 协方差阵的性质

(1) $\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})=\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y})^{\mathrm{T}}$.

(2) 对任给 $k\times m$ 维常数阵 $\boldsymbol{A}$ 和 k 维常向量 $\boldsymbol{b}$ 有

$$\operatorname{cov}(\boldsymbol{AX}+\boldsymbol{b},\boldsymbol{Y}) = \boldsymbol{A}\operatorname{cov}(\boldsymbol{X},\boldsymbol{Y}).$$

(3) $\operatorname{cov}(\boldsymbol{X}+\boldsymbol{Y},\boldsymbol{Z})=\operatorname{cov}(\boldsymbol{X},\boldsymbol{Z})+\operatorname{cov}(\boldsymbol{Y},\boldsymbol{Z})$.

定义 4.5.3 对随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$，$\boldsymbol{X}$ 的**方差阵**(variance matrix)为

$$V(\boldsymbol{X}) = \operatorname{cov}(\boldsymbol{X},\boldsymbol{X}) = (\sigma_{ij})_{n\times n},$$

其中 $\sigma_{ij}=\operatorname{cov}(X_i,X_j)$，$i,j=1,\cdots,n$.

在一些文献中方差阵又称为**方差-协方差阵**(variance-covariance matrix).

性质 4.5.4 方差阵的性质

(1) $V(\boldsymbol{X})$为对称非负定矩阵，即对任给的 n 维列向量 $\boldsymbol{a}$，有 $\boldsymbol{a}^{\mathrm{T}}V(\boldsymbol{X})\boldsymbol{a}\geqslant 0$.

(2) 对任给的 n 维常向量 $\boldsymbol{a}$，有

$$\sigma^2(\boldsymbol{a}^{\mathrm{T}}X) = \boldsymbol{a}^{\mathrm{T}}V(\boldsymbol{X})\boldsymbol{a}.$$

(3) 对任给的 $k\times n$ 维常数阵 $\boldsymbol{A}$，有

$$V(\boldsymbol{AX}) = \boldsymbol{A}V(\boldsymbol{X})\boldsymbol{A}^{\mathrm{T}}.$$

(4) $|V(\boldsymbol{X})|=0$，当且仅当存在 n 维常向量 $\boldsymbol{a}$ 和常数 c，使得 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}=c$.

定义 4.5.5 设随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$，其**相关阵**

(correlation matrix)为

$$\mathrm{cor}(\boldsymbol{X}) = (\rho_{ij})_{n\times n},$$

其中 $\rho_{ij}=\rho(X_i,X_j),i,j=1,\cdots,n.$

性质 4.5.6 相关阵的性质

(1) $\mathrm{cor}(\boldsymbol{X})$为对角线元素均为 1 的对称非负定矩阵.

(2) 设 $\boldsymbol{\Sigma}=(\sigma_{ij})_{n\times n}$为方差阵,$\boldsymbol{D}=\mathrm{diag}\left(\sigma_{11}^{\frac{1}{2}},\cdots,\sigma_{nn}^{\frac{1}{2}}\right)$,则$\mathrm{cor}(X)=\boldsymbol{D}^{-1}\boldsymbol{\Sigma}\boldsymbol{D}^{-1}$.

(3) 设 $\boldsymbol{C}=(\rho_{ij})_{n\times n}$为相关阵,$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 为 $\boldsymbol{C}$ 的特征根,则有

$$\lambda_n \geqslant 0,\ \sum_{i=1}^{n}\lambda_i = n.$$

(4) $\boldsymbol{C}$(相关阵)的行列式 $|\boldsymbol{C}| = \prod_{i=1}^{n}\lambda_i \leqslant 1$, $|\boldsymbol{C}| = 1$,当且仅当 $\boldsymbol{C}=\boldsymbol{I}_n$:($n$ 阶单位阵);$|\boldsymbol{C}|=0$,当且仅当存在 n 维常向量 $\boldsymbol{a}$ 及常数 C,使得 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}=c$,其中 $\boldsymbol{X}$ 为 $\boldsymbol{C}$ 所生成的随机向量.

推论 4.5.7 设 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$ 为 n 维随机向量,且 $\sigma^2(X_i)=1,i=1,\cdots,n$,则 $V(\boldsymbol{X})=\mathrm{cor}(\boldsymbol{X})$.

定理 4.5.8 设 $\boldsymbol{\Sigma}$ 为 n 维随机向量 $\boldsymbol{X}$ 的方差阵,$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0$ 为 $\boldsymbol{\Sigma}$ 的特征根,相应的单位特征向量为 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_n$,则有

$$\lambda_1 = \sigma^2(\boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{X}) = \max_{\|\boldsymbol{a}\|=1}\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}),$$

$$\lambda_2 = \sigma^2(\boldsymbol{\eta}_2^{\mathrm{T}}\boldsymbol{X}) = \max_{\substack{\|\boldsymbol{a}\|=1\\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\eta}_1=0}}\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}),$$

$$\cdots\cdots$$

$$\lambda_k = \sigma^2(\boldsymbol{\eta}_k^{\mathrm{T}}\boldsymbol{X}) = \max_{\substack{\|\boldsymbol{a}\|=1\\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\eta}_1=0,\cdots,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\eta}_{k-1}=0}}\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}),$$

$$\cdots\cdots$$

$$\lambda_n = \sigma^2(\boldsymbol{\eta}_n^{\mathrm{T}}\boldsymbol{X}) = \min_{\|\boldsymbol{a}\|=1}\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}).$$

5 特征函数与母函数

5.1 特征函数

5.1.1 一元特征函数

定义 5.1.1 对随机变量 X 及其分布函数 $F(x)$，相应的**特征函数**(characteristic function)为

$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x),\quad -\infty<t<+\infty,$$

其中 i 为虚数单位，

$$\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x)=\int_{-\infty}^{\infty}\cos(tx)\mathrm{d}F(x)+\mathrm{i}\int_{-\infty}^{\infty}\sin(tx)\mathrm{d}F(x).$$

公式 5.1.2 反转公式(inversion formula)

$$\frac{F(b)+F(b-)}{2}-\frac{F(a)+F(a-)}{2}$$

$$=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}ta}-\mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\varphi(t)\mathrm{d}t.$$

对任意的 $-\infty<a<b<+\infty$. 特别当 a,b 为 $F(x)$ 的连续点时，上式左端为 $(F(b)-F(a))/2$.

定理 5.1.3 唯一性定理 分布函数与特征函数相互唯一确定.

定理 5.1.4 若 $\int_{-\infty}^{\infty}|\varphi(t)|\mathrm{d}t<\infty$，则相应的分布函数 $F(x)$ 的导数存在并连续，且

$$F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t.$$

性质 5.1.5 特征函数的性质

(1) $|\varphi(t)|\leqslant|\varphi(0)|=1$;

(2) $\varphi(-t)=\overline{\varphi(t)}$;

(3) $\varphi(t)$在整个 $\mathbf{R}$ 上一致连续;

(4) $\varphi(t)$为实值函数,当且仅当相应的分布函数关于原点对称;

(5) 设 X 的特征函数为 $\varphi(t)$,则 $aX+b$ 的特征函数为 $e^{ibt}\varphi(at)$;

(6) 对任意实数 $t_1,\cdots,t_n$ 及复数 $\lambda_1,\cdots,\lambda_n,n=1,2,\cdots$,有

$$\sum_{k=1}^{n}\sum_{l=1}^{n}\varphi(t_k-t_l)\lambda_k\overline{\lambda_l}\geqslant 0;$$

(7) 设 $E(X^m)$存在,则 $\varphi(t)$的 m 阶导数存在,且

$$\varphi^{(k)}(t)|_{t=0}=i^kE(X^k),k=1,2,\cdots,m;$$

(8) 设 $X_1,\cdots,X_n$ 相互独立,$X=\sum_1^n X_i$,则

$$\varphi_X(t)=\prod_{i=1}^{n}\varphi_{X_i}(t).$$

定理 5.1.6 设随机变量 X 的特征函数 $\varphi(t)$在某个 $t_0\neq 0$ 点模为 $1(|\varphi(t_0)|=1,t_0\neq 0)$,则 X 为取值于$\{(2k\pi+a)/t_0,k=0,\pm1,\pm2,\cdots\}$的离散型随机变量,其中 $a=-\mathrm{i}\log[\varphi(t_0)]$;反之亦真.

定义 5.1.7 设特征函数 $\varphi(t)$满足

$$\log\varphi(t)=\sum_{j=1}^{k}\kappa_j(\mathrm{i}t)^j/j!+0(t^k),当 t\to 0,$$

则称 κ_j 为 $\varphi(t)$的 j 阶**半不变量**(semi-invariant)或**累积量**(cumulant),$j=1,\cdots,k$. $\log\varphi(t)$又称为**累积量生成函数**(cumulant generation function).

推论 5.1.8 记随机变量 X 的 j 阶半不变量为 $\kappa_j(X)$,则有

(1) $\kappa_j(X+b)=\kappa_j(X)$;

(2) $\kappa_j(aX)=a^j\kappa_j(X)$;

(3) 若 $X_1,\cdots,X_n$ 相互独立，则

$$\kappa_j\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\kappa_j(X_i).$$

前 5 阶半不变量与矩的变换公式

$$\kappa_1=E(X)=\alpha_1,\quad \kappa_2=\mu_2,\quad \kappa_3=\mu_3,$$

$$\kappa_4=\mu_4-3\mu_2^2,\quad \kappa_5=\mu_5-10\mu_3\mu_2,$$

其中 μ_i 为 i 阶中心矩，α_1 为一阶原点矩.

5.1.2 多元特征函数

定义 5.1.9 设随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的分布函数为 $F(x_1,x_2,\cdots,x_n)$，其特征函数为

$$\varphi(t_1,\cdots t_n)=E\left[\mathrm{e}^{\mathrm{i}\sum_{1}^{n}t_jX_j}\right]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\sum_{1}^{n}t_jx_j}\mathrm{d}F(x_1,\cdots,x_n).$$

定理 5.1.10 唯一性定理 多元分布函数与多元特征函数相互唯一确定.

性质 5.1.11 多元特征函数的性质

(1) $|\varphi(t_1,\cdots,t_n)|\leqslant\varphi(0,\cdots,0)=1$；

(2) $\varphi(-t_1,\cdots,-t_n)=\overline{\varphi(t_1,\cdots,t_n)}$；

(3) 若 $(X_1,\cdots,X_n)$ 的特征函数为 $\varphi(t_1,\cdots,t_n)$，则对 $k<n$，$\varphi(t_1,\cdots,t_k,0,\cdots,0)$ 为 $(X_1,\cdots,X_k)$ 的特征函数；

(4) $\sum_{k=1}^{n}a_kX_k$ 的特征函数为 $\varphi(a_1t_1,\cdots,a_nt_n)$；

(5) $X_1,\cdots,X_n$ 相互独立的充分必要条件为

$$\varphi(t_1,\cdots,t_n)=\varphi_{X_1}(t_1)\cdots\varphi_{X_n}(t_n);$$

(6) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$，$\boldsymbol{Y}=(Y_1,\cdots,Y_m)$，则 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立的充分必要条件为

$$\varphi(t_1,\cdots,t_n,t_{n+1},\cdots,t_{n+m})=\varphi_X(t_1,\cdots,t_n)\varphi_Y(t_{n+1},\cdots,t_{n+m}),$$

其中 φ,φ_X 和 φ_Y 分别为 $(\boldsymbol{X},\boldsymbol{Y})$，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的特征函数；

(7) 若 $E(X_1^{k_1}X_2^{k_2}\cdots X_n^{k_n})$存在，则

$$E(X_1^{k_1}X_2^{k_2}\cdots X_n^{k_n}) = \mathrm{i}^{\left(\sum\limits_{j=1}^{n}k_j\right)}\left[\frac{\partial^{\sum k_j}\varphi}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}}\right]_{t_1=\cdots=t_n=0}.$$

5.2 母函数

5.2.1 概率母函数

本节只考虑取非负整数值的随机变量，称为整型随机变量.

定义 5.2.1 设 X 有分布列 $p_k=P(X=k),k=0,1,2,\cdots,X$ 的**概率母函数**(probability generating function)为

$$g(t) = \sum_{k=0}^{\infty}p_kt^k = E(t^X), \quad -1\leqslant t\leqslant 1.$$

以下简称概率母函数为母函数.

定理 5.2.2 母函数在$[-1,1]$上一定存在，且与概率分布列相互唯一确定.

性质 5.2.3 母函数的性质

(1) $g(0)=p_0,g^{(k)}(0)=k!\,p_k,k=1,2,\cdots$;

(2) 若 $E(X^k)$存在，则 $g^{(k)}(1)=E[X(X-1)\cdots(X-k+1)]$;

(3) 设 $X_1,\cdots,X_n$ 为相互独立的随机变量，有母函数 $g_1(t)$，$\cdots,g_n(t)$，则 $X=\sum\limits_{i=1}^{n}X_i$ 有母函数

$$g(t) = \prod_{i=1}^{n}g_i(t);$$

(4) 设 $X_1,\cdots,X_n,\cdots$为独立同分布随机变量列，有相同的母函数 $g_1(t)$，N 为取非负整值的随机变量，有母函数 $g_2(t)$，且与$\{X_i,i=1,2,\cdots\}$相互独立，定义随机变量

$$X = \sum_{i=0}^{N}X_i = \begin{cases}\sum\limits_{i=1}^{n}X_i, & 当\ N=n,n=1,2,\cdots,\\ 0, & 当\ N=0,\end{cases}$$

则 X 有母函数 $g(t)=g_2(g_1(t))$.

例 5.2.4 设 $X_1,X_2,\cdots$为独立随机变量列. $P(X_i=1)=p$, $P(X_i=0)=1-p$, N 与$\{X_1,X_2,\cdots\}$独立,服从泊松分布 $P(\lambda)$,则 $X_1,X_2,\cdots$有相同的母函数 $1-p+pt$, N 有母函数 $e^{-\lambda(1-t)}$,所以 $X=\sum_{i=0}^{N}X_i$ 有母函数

$$e^{-\lambda[1-(1-p+pt)]}=e^{-\lambda p(1-t)}.$$

定义 5.2.5 对整型随机变量 X,记 $f_k=P(X>k)$, $k=0,1,2,\cdots$,若和式 $\sum_{k=0}^{\infty}f_kt^k$ 在区间 $-t_0<t<t_0$ 上收敛,则称在$(-t_0,t_0)$上的函数

$$q(t)=\sum_{k=0}^{\infty}f_kt^k$$

为 X 的**尾概率母函数**(generating function of tail probability).

定理 5.2.6 尾概率母函数在区间$(-1,1)$上一定存在,在区间$[-1,1]$上存在的充分必要条件为 $E(X)$存在且有限.

性质 5.2.7 尾概率母函数的性质

(1) $q(1)=E(X)$;

(2) 若 $g(t)$为概率母函数,则

$$q(t)=\frac{1-g(t)}{1-t}.$$

5.2.2 矩母函数

定义 5.2.8 对随机变量 X 及其分布函数 $F(x)$,若积分 $\int_{-\infty}^{\infty}e^{tx}dF(x)$ 在某一区间(t_1,t_2)上存在且有限,则定义区间(t_1,t_2)上的函数

$$m(t)=\int_{-\infty}^{\infty}e^{tx}dF(x)=E(e^{tX})$$

为 X 的**矩母函数**(moment generating function).

性质 5.2.9 矩母函数的性质

(1) $m(0)=1$;

(2) 若 $m(t)$在包含原点的开区间(t_1,t_2)上存在,则在此区间上的各阶导数都存在,且

$$m^{(k)}(0) = E(X^k), k = 1,2,\cdots;$$

(3) 设 X 的矩母函数为 $m(t)$,则 $aX+b$ 的矩母函数为

$$e^{bt}m(at);$$

(4) 设 $X_1,\cdots,X_n$ 相互独立且分别有矩母函数 $m_1(t),\cdots,m_n(t)$,则 $X = \sum_{i=1}^{n} X_i$ 有矩母函数 $\prod_{i=1}^{n} m_i(t)$.

6 常用分布

6.1 与伯努利试验有关的离散分布

伯努利(Bernoulli)试验序列是一种简单而有广泛应用的概率模型.由之产生出一系列常用的概率分布.

6.1.1 0-1分布与两点分布

定义 6.1.1 设随机变量 X 只取 0 和 1 两个值,记 $p=P(X=1)$,$q=1-p=P(X=0)$,称 X 服从 0-1 分布.

做一次"成功"概率为 p 的伯努利试验,定义随机变量

$$X=\begin{cases}1, & \text{当试验“成功”},\\ 0, & \text{当试验“失败”}.\end{cases}$$

则 X 服从 0-1 分布.

定义 6.1.2 设随机变量 Y 只取 a 和 b 两个值,记 $p=P(Y=b)$,$q=1-p=P(Y=a)$,则称 Y 服从 **a-b 两点分布**.

若 X 服从 0-1 分布,则对任意两个实数 a 和 b,$(b-a)X+a$ 服从 a-b 两点分布.反之,若 Y 服从 a-b 两点分布,则 $(Y-a)/(b-a)$ 服从 0-1 分布.

定义 6.1.3 0-1 分布的特征

(1) $E(X)=p$,$\sigma^2(X)=pq$.

(2) 概率母函数　$g(t)=pt+q$,

矩母函数　$m(t)=pe^t+q$,

特征函数　$\varphi(t)=pe^{it}+q$.

6.1.2 二项分布

定义 6.1.4 设随机变量取值于$\{0,1,\cdots,n\}$,满足

$$P(X=k)=b(k;n,p)=\binom{n}{k}p^kq^{n-k},k=0,1,\cdots,n,$$

其中 $p\in(0,1)$,$q=1-p$,称 X 服从参数为 n,p 的**二项分布**(binomial distribution),记作 $X\sim B(n,p)$.

二项分布是 n 重伯努利试验中"成功次数"的分布.又设 $X_1,\cdots,X_n$ 为独立有相同 0-1 分布的随机变量,则 $X=\sum_{i=1}^{n}X_i\sim B(n,p)$.

性质 6.1.5 二项分布的特征与性质

(1) $E(X)=np$,$\sigma^2(X)=npq$.

(2) 概率母函数 $g(t)=(pt+q)^n$,

矩母函数 $m(t)=(pe^t+q)^n$,

特征函数 $\varphi(t)=(pe^{it}+q)^n$.

(3) $b(k;n,p)=b(n-k;n,1-p)$,

$$\frac{b(k+1;n,p)}{b(k;n,p)}=\frac{n-k+1}{k+1}\frac{p}{q}.$$

二项分布列的示意如图 6.1

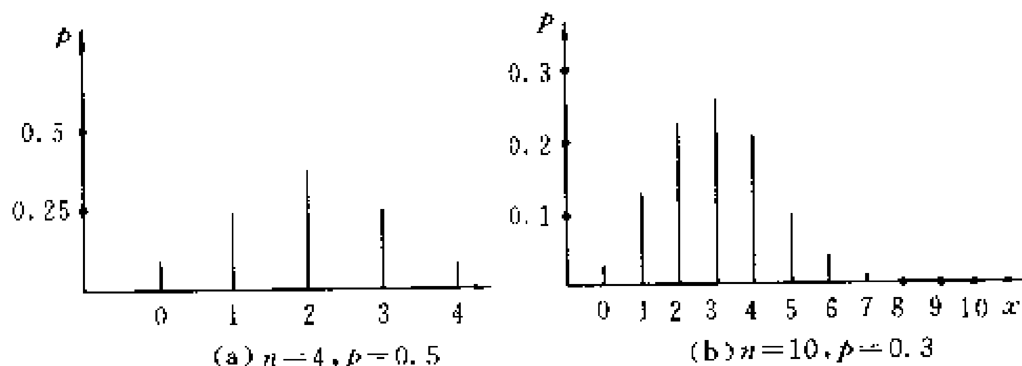


图 6.1

二项分布概率的计算比较麻烦. 当 npq 较大时(例如 $0.2\leqslant p\leqslant 0.8, npq\geqslant 10$), 可用正态逼近.

$$\sum_{i=0}^{k} b(i;n,p) \approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{k-np}{\sqrt{npq}}} e^{-\frac{x^2}{2}}dx = \Phi\left(\frac{k-np}{\sqrt{npq}}\right),$$

其中 $\Phi(\cdot)$ 为标准正态分布函数. 当 p 较小因而 np 不大时(例如 $p\leqslant 0.1, np\approx 1$)可用泊松逼近.

$$b(k;n,p) \approx \frac{\lambda^k}{k!}e^{-\lambda} = p(k;\lambda), \text{其中 } \lambda = np.$$

定理 6.1.6 设 X_1 与 X_2 为两个相互独立的非负整值随机变量, 则 $X_1+X_2\sim B(n,p)$, 当且仅当 $X_1\sim B(n_1,p)$, $X_2\sim B(n_2,p)$, 且 $n_1+n_2=n$.

6.1.3 几何分布

定义 6.1.7 设随机变量 X 取正整值, 满足

$$P(X=k) = pq^{k-1} \quad k=1,2,\cdots.$$

其中 $p\in(0,1)$, $q=1-p$, 称 X 服从**几何分布**(geometric distribution).

进行"成功"概率为 p 的伯努利试验, 直到首次"成功"出现为止, 用 X 记所需的试验次数, 则 $X=k$ 意味着前 $k-1$ 次"失败"而第 k 次"成功", 相应的概率为 $q^{k-1}p$, 即几何分布.

性质 6.1.8 几何分布的特征

(1) $E(X)=\dfrac{1}{p}$, $\sigma^2(X)=q/p^2$.

(2) 概率母函数 $g(t)=\dfrac{pt}{1-qt}$, $-\dfrac{1}{q}<t<\dfrac{1}{q}$;

矩母函数 $m(t)=\dfrac{pe^t}{1-qe^t}$ $\quad -\infty<t<-\log q$;

特征函数 $\varphi(t)=\dfrac{pe^{it}}{1-qe^{it}}$.

几何分布具有无记忆性, 即满足

$$P(X=k+r \mid X>k)=P(X=r).$$

6.1.4 帕斯卡分布与负二项分布

定义 6.1.9 设随机变量 X 取值于 $\{r,r+1,\cdots\}$,

$$P(X=k)=\binom{k-1}{r-1}p^r q^{k-r},\quad k=r,r+1,\cdots.$$

则称 X 服从**帕斯卡(Pascal)分布**.

进行"成功"概率为 p 的伯努利试验,直到第 r 次成功出现为止,记 X 为试验次数,则 X 服从帕斯卡分布.又设 $X_1,X_2,\cdots,X_r$ 为相互独立具相同几何分布的随机变量,则 $X=\sum_{i=1}^{r}X_i$ 服从帕斯卡分布.

性质 6.1.10 帕斯卡分布的特征

(1) $E(X)=r/p,\sigma^2(X)=rq/p^2$;

(2) 概率母函数 $g(t)=\left(\dfrac{pt}{1-qt}\right)^r,-\dfrac{1}{q}<t<\dfrac{1}{q}$,

矩母函数 $m(t)=\left(\dfrac{pe^t}{1-qe^t}\right)^r,-\infty<t<-\log q$,

特征函数 $\varphi(t)=\left(\dfrac{pe^{it}}{1-qe^{it}}\right)^r$.

定义 6.1.11 设随机变量 Y 取非负整值,

$$P(Y=k)=\binom{-r}{k}p^r(-q)^k,\quad k=0,1,2,\cdots.$$

其中 $r>0,p\in(0,1),q=1-p$,称 Y 服从**负二项分布**(negative binomial distribution).

推论 6.1.12 设 r 为正整数,$p\in(0,1)$,若 X 服从参数为 r,p 的帕斯卡分布,则 $X-r$ 服从负二项分布.

对一般的负二项分布,r 不必是正整数.

性质 6.1.13 负二项分布的特征

(1) $E(Y)=\dfrac{r}{p}-r=r\dfrac{q}{p},\sigma^2(Y)=rq/p^2$;

(2) 概率母函数 $g(t)=\left(\frac{p}{1-qt}\right)^{r}, -\frac{1}{q}<t<\frac{1}{q}$,

矩母函数 $m(t)=\left(\frac{p}{1-qe^{t}}\right)^{r}, -\infty<t<\log q$,

特征函数 $\varphi(t)=\left(\frac{p}{1-qe^{it}}\right)^{r}$.

6.2 泊松分布及有关的分布

6.2.1 泊松分布

定义 6.2.1 设 X 为非负整数值随机变量,

$$P(X=k)=p(k;\lambda)=\frac{\lambda^{k}}{k!}e^{-\lambda},\quad k=0,1,2,\cdots.$$

其中 $\lambda>0$ 为常数,称 X 服从参数为 λ 的**泊松**(Poisson)**分布**,记为 $X\sim P(\lambda)$.

泊松分布有广泛应用. 例如,在单位时间内放射性物质放射出的 α 粒子数. 路口通过的车辆数,服务台到达的顾客数等,都可以用泊松分布来刻画. 泊松分布产生的机理是基于一种平稳的独立增量过程. 假设一块放射性物质放射出的 α 粒子数在一段时间间隔内具有如下的性质:

性质 6.2.2 (1) 在任一时间间隔 Δt 内放射的 α 粒子数的分布与此间隔之外的 α 粒子数的分布无关(独立增量性);

(2) 在 Δt 内放射的 α 粒子数的分布仅与 Δt 的长度有关,而与 Δt 的位置无关(平稳性);

(3) 记 $P_{\Delta t}(k)$ 为 Δt 时间间隔内放射 k 个 α 粒子的概率,$k=0,1,2,\cdots$,有

$$P_{\Delta t}(0)<1;P_{\Delta t}(1)/\Delta t\to\lambda>0,\quad 当\ \Delta t\to 0;$$

及

$$\sum_{k=2}^{\infty}P_{\Delta t}(k)/P_{\Delta t}(1)=o(1),\quad 当\ \Delta t\to 0.$$

则在单位时间内放射出的 α 粒子数服从 $P(\lambda)$分布；而在 t 时间间隔内放射出的 α 粒子数服从 $P(\lambda t)$分布.

性质 6.2.3 泊松分布的特征和性质

(1) $E(X)=\lambda, \sigma^2(X)=\lambda$;

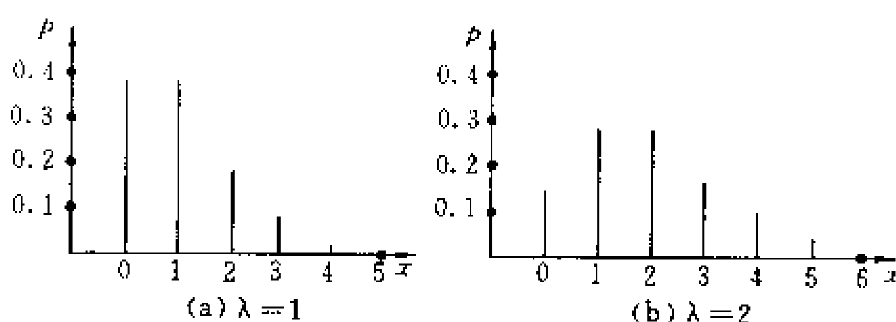
(2) 概率母函数 $g(t)=\mathrm{e}^{-\lambda(1-t)}$,

矩母函数 $m(t)=\mathrm{e}^{-\lambda(1-\mathrm{e}^t)}$,

特征函数 $\varphi(t)=\mathrm{e}^{-\lambda(1-\mathrm{e}^{it})}$;

(3) $p(k+1;\lambda)/p(k;\lambda)=\dfrac{\lambda}{k+1}$.

泊松分布列的示意如图 6.2



(a) λ=1　　(b) λ=2

图 6.2

计算泊松分布概率 $p(k;\lambda)$可利用斯特林(Stirling)公式：

$$k! \approx \sqrt{2\pi} k^{k+\frac{1}{2}} \mathrm{e}^{-k},$$

得到 $p(k;\lambda)$的近似公式

$$p(k;\lambda) \approx (2\pi k)^{-\frac{1}{2}} \mathrm{e}^{-\lambda} \left(\frac{\lambda \mathrm{e}}{k}\right)^k.$$

定理 6.2.4 设 X_1, X_2 为独立非负整值随机变量，则 $X_1+X_2 \sim P(\lambda)$的充分必要条件是

$$X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2), \text{且 } \lambda_1 + \lambda_2 = \lambda.$$

6.2.2 复合泊松分布

定义 6.2.5 设 $X_1, X_2, \cdots$为一串独立同分布的非负整值随

机变量列，N 为与$\{X_1,X_2,\cdots\}$独立的泊松分布随机变量，则 $Y=\sum_{i=1}^{N}X_i$ 的分布称为**复合泊松分布**.

推论 6.2.6 设 $Y=\sum_{i=1}^{N}X_i$ 具复合泊松分布，则 Y 有概率母函数

$$G(t)=\mathrm{e}^{-\lambda(1-g(t))}.$$

其中 $g(t)$为 X_1 的概率母函数，λ 为泊松分布的参数.

定理 6.2.7 设 X 为一非负整值随机变量，则对任给正整数 n，X 可表为 n 个独立同分布随机变量之和的充分必要条件为 X 服从复合泊松分布.

上述定理中所刻画的复合泊松分布的特性称为**无穷可分性**(infinite divisability). 复合泊松分布是一般无穷可分分布族的一个特例.

例 6.2.8 设 $X_1,X_2,\cdots$为独立同分布 0-1 变量列，N 为与$\{X_1,X_2,\cdots\}$独立的 $P(\lambda)$分布随机变量，则 $Y=\sum_{i=1}^{N}X_i\sim P(\lambda p)$.

例 6.2.9 负二项分布的概率母函数为

$$g(t)=\left(\frac{p}{1-qt}\right)^r=\mathrm{e}^{-\lambda(1-f(t))},$$

其中 $\lambda=-r\log p$，$f(t)=\dfrac{\log(1-qt)}{\log p}$为对数分布的概率母函数. 因此负二项分布为复合泊松分布.

6.3 其它离散分布

6.3.1 退化分布

定义 6.3.1 设随机变量 X 只取一个值：$P(X=C)=1$，则称 X 服从**退化分布**(degenerate distribution).

实际上退化分布就是把常数作为随机变量来处理,因而带来很大的方便.

推论 6.3.2 设 X 服从退化分布,则 X 与任一个或任一组随机变量独立.

性质 6.3.3 退化分布的特征

(1) $E(X)=C,\sigma^2(X)=0$;

(2) 矩母函数 $m(t)=e^{ct}$,

特征函数 $\varphi(t)=e^{ict}$.

定理 6.3.4 X 服从退化分布当且仅当 $\sigma^2(X)=0$.

6.3.2 离散均匀分布

定义 6.3.5 设随机变量 X 取值于 $\{0,1,2,\cdots,n\}$,且 $P(X=0)=P(X=1)=\cdots=P(X=n)=\frac{1}{n+1}$,则称 X 服从**离散均匀分布**(discrete uniform distribution).

性质 6.3.6 离散均匀分布的特征

(1) $E(X)=\frac{n}{2},\sigma^2(X)=\frac{n(n+2)}{12}$;

(2) 概率母函数 $g(t)=\frac{1-t^{n+1}}{(n+1)(1-t)}$,

矩母函数 $m(t)=\frac{1-e^{(n+1)t}}{(n+1)(1-e^t)}$,

特征函数 $\varphi(t)=\frac{1-e^{i(n+1)t}}{(n+1)(1-e^{it})}$.

利用离散均匀分布可产生十进制的随机数.设 $X_1,X_2,\cdots,X_k$ 为独立的取值于 $\{0,1,\cdots,9\}$ 的离散均匀分布随机变量,则 $Y=0.X_1,X_2,\cdots,X_k$ 构成 $[0,1)$ 区间上的 k 位随机数.

6.3.3 超几何分布

定义 6.3.7 设随机变量 X 取值于 $\{0,1,2,\cdots n\}$,

$$P(X=k)=h(k)=\binom{N}{k}\binom{M}{n-k}\Big/\binom{M+N}{n},$$
$$k=0,1,\cdots,n,$$
其中 M,N 均为正整数且 $n\leqslant N+M$，则称 X 服从**超几何分布**(hypergeometric distribution)，记作 $X\sim H(N,M)$.

超几何分布是由无放回抽样产生的分布.

性质 6.3.8 超几何分布的特征

(1) $E(X)=n\dfrac{N}{N+M}$，$\sigma^2(X)=n\dfrac{NM}{(N+M)^2}\dfrac{N+M-n}{N+M-1}$；

(2) 各种母函数均无简单形式.

超几何分布用于产品质量的抽样检验，当产品数目远远大于抽样个数时，可采用二项逼近.

$$h(k)\to b(k;n\cdot p),\text{当 } N+M\to\infty,\frac{N}{N+M}\to p.$$

6.3.4 对数分布

定义 6.3.9 设随机变量 X 取正数整值，

$$P(X=k)=-q^k/(k\log p),k=1,2,\cdots.$$

其中 $p\in(0,1)$ 为常数，$q=1-p$，则称 X 服从**对数分布**(logrithm distribution).

性质 6.3.10 对数分布的特征

(1) $E(X)=-q/(p\log p)$，$\sigma^2(X)=\dfrac{-q}{p^2\log p}\left[1-\dfrac{q}{(p\log p)^2}\right]$；

(2) 概率母函数 $g(t)=\log(1-qt)/\log p,-q^{-1}<t<q^{-1}$，

矩母函数 $m(t)=\log(1-qe^t)/\log p,-\infty<t<-\log q$，

特征函数 $\varphi(t)=\log(1-qe^{it})/\log p$.

定理 6.3.11 由对数分布产生的复合泊松分布为负二项分布.

6.4 正态分布与有关的分布

6.4.1 正态分布

定义 6.4.1 设连续型随机变量 X 有密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad -\infty<x<+\infty,$$

其中 μ,σ 均为常数，$-\infty<\mu<+\infty,\sigma>0$，称 X 服从**正态分布**(normal distribution)，又称为**高斯**(Gauss)**分布**，记作 $X\sim N(\mu,\sigma^2)$.

正态分布的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(u-\mu)^2}{2\sigma^2}}\mathrm{d}u,\quad -\infty<x<+\infty.$$

正态分布是概率论和数理统计中应用最为广泛的分布. 由中心极限定理(见第 10 章)知，当一个随机变量可表示为许多个随机变量之和，其中每个随机变量对总和的影响都不起决定性作用时，和的分布可近似地看成是正态的. 例如，测量的误差，人的身高、体重，炮弹弹着点的分布，农作物的产量等都可以用正态分布来近似地刻画. 另一方面，许多常用分布(例如二项、泊松等分布)，也可在一定条件下用正态分布来近似.

性质 6.4.2 正态分布的特征

(1) $E(X)=\mu,\ \sigma^2(X)=\sigma^2$;

(2) 矩母函数 $m(t)=e^{\mu t+\frac{\sigma^2t^2}{2}}\quad -\infty<t<+\infty$,

特征函数 $\varphi(t)=e^{i\mu t-\frac{\sigma^2t^2}{2}}$;

(3) 密度关于 $x=\mu$ 点对称，μ 也是中位数和众数；

(4) k 阶中心矩

$$\mu_k=E(X-\mu)^k=\begin{cases}0 & \text{当 } k \text{ 为奇数;}\\ (k-1)!!\sigma^k, & \text{当 } k \text{ 为偶数,}\end{cases}$$

其中$(k-1)!!=(k-1)(k-3)\cdots 3\cdot 1$.

正态密度图象示意如图 6.3.

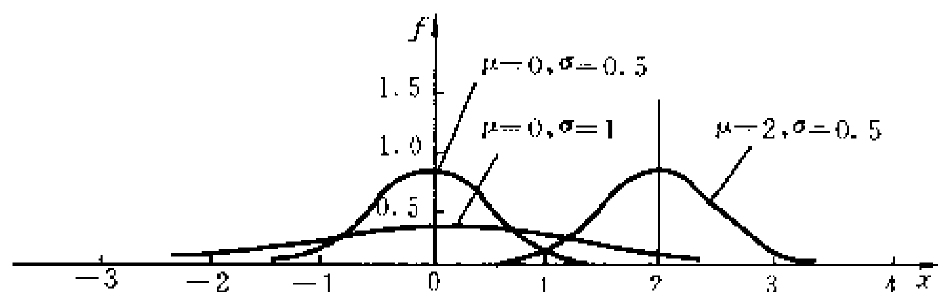


图 6.3

由图 6.3 可见,密度曲线关于 $x=\mu$ 点对称,其形状由 σ 的大小所决定.

定理 6.4.3 克拉默-列维(Cramer-Lévy)定理 设 X_1,X_2 为独立随机变量,则 $X_1+X_2\sim N(\mu,\sigma^2)$,当且仅当 $X_1\sim N(\mu_1,\sigma_1^2)$,$X_2\sim N(\mu_2,\sigma_2^2)$,且 $\mu_1+\mu_2=\mu,\sigma_1^2+\sigma_2^2=\sigma^2$.

定理 6.4.4 设 $X\sim N(\mu,\sigma^2)$,则对常数 $a\neq 0$ 及 $b,aX+b\sim N(a\mu+b,a^2\sigma^2)$.

利用此定理,在计算一般正态分布的概率时,可将其转化为一种特殊的正态分布进行计算.

定义 6.4.5 设 $X\sim N(0,1)$,**称 X 服从标准正态分布**(standard normal distribution).

标准正态分布的密度和分布函数分别记为 $\varphi(x)$和 $\Phi(x)$:

(1) $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\quad -\infty<x<\infty$,

(2) $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\quad -\infty<x<\infty$.

在标准正态分布表中,给出了对应于 $z\geqslant 0$ 的 $\Phi(z)$值.表的最左边一列给出了 $z=0.0\sim 3.0$ 的值,表的最上面一行给出了 z 的第二位小数.表中之值为 $\Phi(z)$的值,例如要查 $\Phi(1.32)$,则在左边

一列找到 1.3，在 1.3 所在的一行找对应最上面一行为 2 的列，得到 0.9066，即为 $\Phi(1.32)$之值. 此表也可由已知 $\Phi(z)$之值，来查找 z. 当 $z<0$ 时，利用公式

$$\Phi(z) = 1 - \Phi(-z).$$

便可得到 $\Phi(z)$之值. 例如 $\Phi(-1.32)=1-\Phi(1.32)=1-0.9066=0.0934$.

对一般正态分布，利用下述推论将其转化为标准正态分布进行计算.

推论 6.4.6 设 $X\sim N(\mu,\sigma^2)$，则 $(X-\mu)/\sigma\sim N(0,1)$.

例 6.4.7 设 $X\sim N(1.40,0.0025)$，求 $P(1.35<X<1.45)$ 及 X 的上 0.05 分位数.

解 由 $\mu=1.40,\sigma=0.05$ 知 $\dfrac{X-1.40}{0.05}\sim N(0,1)$.

$$P(X<1.45)=P\left(\frac{X-1.40}{0.05}<1.00\right)=\Phi(1.00)=0.8413,$$

$$P(X<1.35)=P\left(\frac{X-1.40}{0.05}<-1.00\right)=\Phi(-1.00)$$
$$=1-\Phi(1.00)=0.1587.$$

于是

$$P(1.35<X<1.45)=P(X<1.45)-P(X<1.35)$$
$$=0.8413-0.1587=0.6826.$$

用 $z_{0.05}$记 $N(0,1)$的上 0.05 分位数，有 $\Phi(z_{0.05})=1-0.05=0.95$，于是 $z_{0.05}=1.65$，于是 x 的上 0.05 分位数 $x_{0.05}$为

$$x_{0.05} = \mu + z_{0.05}\sigma = 1.40 + 1.65 \times 0.05 = 1.4825.$$

6.4.2 χ^2 分布

定义 6.4.8 设 $X_1,X_2,\cdots,X_n$ 为相互独立，同 $N(0,1)$分布的随机变量，定义 $Q = \sum_{i=1}^{n} X_i^2$，则 Q 的分布称为具自由度 n 的 χ^2

分布(chi squared distribution),记作 $Q\sim\chi^2(n)$.

$\chi^2(n)$的概率密度函数为

$$f(x;n)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

其中 $\Gamma\left(\dfrac{n}{2}\right)$ 为伽玛(Gama)函数:

$$\Gamma\left(\frac{n}{2}\right)=\int_0^{\infty}x^{\frac{n}{2}-1}\mathrm{e}^{-x}\mathrm{d}x$$

$\chi^2(n)$的密度函数图象示意图 6.4.

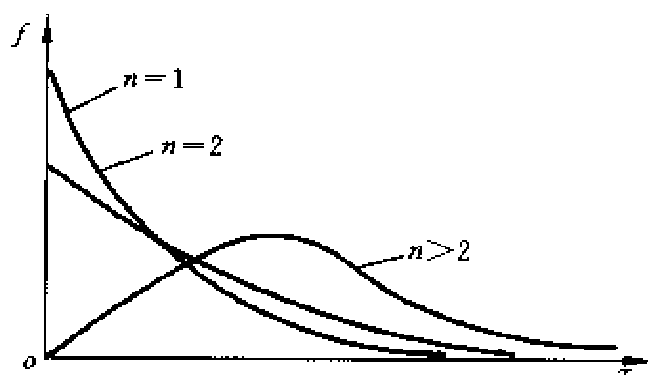


图 6.4

χ^2 分布在数理统计中广泛地用于方差的估计和检验.

性质 6.4.9 $\chi^2(n)$的特征

(1) $E(Q)=n,\sigma^2(Q)=2n$;

(2) 矩母函数 $m(t)=(1-2t)^{\frac{n}{2}}\quad -\infty<t<1$,

特征函数 $\varphi(t)=(1-2\mathrm{i}t)^{-\frac{n}{2}}$.

定理 6.4.10 设 Q_1,Q_2 为相互独立,分别服从 $\chi^2(n_1)$和$\chi^2(n_2)$的随机变量,则 $Q_1+Q_2\sim\chi^2(n_1+n_2)$.

定义 6.4.11 设 $X_1,X_2,\cdots,X_n$ 相互独立,$X_i\sim N(\mu_i,1)$,$i=1,\cdots,n$,且 $\delta=\sqrt{\sum\limits_{i=1}^{n}\mu_i^2}\neq 0$,则 $Q=\sum\limits_{i=1}^{n}X_i^2$ 的分布称为具自由度

n 和非中心参数 δ 的**非中心 χ^2 分布**(noncentral chi squared distribution),记作 $Q\sim\chi^2(n,\delta)$.

由定义可见 $\chi^2(n)$ 是 $\chi^2(n,\delta)$ 当 $\delta=0$ 时的特例,因此又称 $\chi^2(n)$ 为**中心 χ^2 分布**.

$\chi^2(n,\delta)$ 的密度函数为

$$f(x;n,\delta)=\begin{cases}\mathrm{e}^{-\frac{\delta^2+x}{2}}\sum_{i=0}^{\infty}\frac{1}{i!}\left(\frac{\delta^2}{2}\right)^i\left[\frac{x^{i+\frac{n}{2}-1}}{2^{i+\frac{n}{2}}\Gamma\left(i+\frac{n}{2}\right)}\right], & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

性质 6.4.12 $\chi^2(n,\delta)$ 的特征

(1) $E(Q)=n+\delta,\sigma^2(Q)=2n$;

(2) 特征函数 $\varphi(t)=(1-2\mathrm{i}t)^{\frac{n}{2}}\exp\left\{\frac{\mathrm{i}t\delta^2}{1-2\mathrm{i}t}\right\}$.

推论 6.4.13 设 $Q_1\sim\chi^2(n_1,\delta_1),Q_2\sim\chi^2(n_2,\delta_2)$,且 Q_1 与 Q_2 相互独立,则 $Q_1+Q_2\sim\chi^2(n,\delta)$,其中 $n=n_1+n_2,\delta=(\delta_1^2+\delta_2^2)^{\frac{1}{2}}$.

χ^2 分布与独立同方差的正态随机变量的一类二次型有密切联系.以下总设 $X_1,X_2,\cdots,X_n$ 为独立随机变量,$X_i\sim N(\mu_i,\sigma^2)$,$i=1,\cdots,n$. 记列随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$,列常数向量 $\boldsymbol{M}=(\mu_1,\cdots,\mu_n)^{\mathrm{T}}$.

定理 6.4.14 设 $\boldsymbol{A}$ 为 n 阶对称阵,$Q=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}/\sigma^2$,则 $Q\sim\chi^2(k,\delta)$,当且仅当 $\boldsymbol{A}$ 为幂等阵,即 $\boldsymbol{A}^2=\boldsymbol{A}$,此时 $k=\mathrm{rank}(\boldsymbol{A})=\mathrm{trace}(\boldsymbol{A}),\delta=\sigma\boldsymbol{M}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{M}^{\frac{1}{2}}$.

定理 6.4.15 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为两个对称幂等阵,$Q_1=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}/\sigma^2$ 和 $Q_2=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{X}/\sigma^2$ 均服从 χ^2 分布(不必是中心的),且 $\boldsymbol{A}-\boldsymbol{B}$ 为非零半正定阵,则 $(Q_1-Q_2)/\sigma^2$ 也服从 χ^2 分布,自由度为 $\mathrm{rank}(\boldsymbol{A})-\mathrm{rank}(\boldsymbol{B})$,且 $\boldsymbol{B}(\boldsymbol{A}-\boldsymbol{B})=0$.

定理 6.4.16 柯赫兰(Cochran)定理 设 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_m$ 为 m 个对称阵,且 $\sum_{i=1}^{m}\boldsymbol{A}_i=\boldsymbol{I}_n$:单位阵. 记 $Q_i=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}_i\boldsymbol{X}/\sigma^2,i=1,\cdots,m$.

则 $Q_1,\cdots,Q_m$ 为相互独立的 χ^2 分布的充分必要条件为 $\sum_{i=1}^{m}\text{rank}(A_i)=n$. 此时 $Q_i\sim\chi^2(n_i,\delta_i)$, $n_i=\text{rank}(A_i)$, $\delta_i^2=\sigma^2 M' A_i M$, $i=1,\cdots,m$.

定理 6.4.17 设 $X_1,X_2,\cdots,X_n$ 独立,同 $N(\mu,\sigma^2)$分布,记 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$, 则

(1) $\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$,

(2) $(n-1)S^2/\sigma^2\sim\chi^2(n-1)$,

(3) $\overline{X}$ 与 S^2 相互独立.

以上几个结论在数理统计中有重要的应用.

对中心 χ^2 分布,常要查表求上 α 分位数,α 常取 0.9,0.95,0.99 或 0.1,0.05,0.01 等值. 在附表 4(χ^2 分布表)中,最上面一行是 α 值,最左边一列是自由度 n,表中是 $\chi^2(n)$的上 α 分位数(记作 $\chi_\alpha^2(n)$). 例如,查表可得:$\chi_{0.975}^2(10)=3.247$, $\chi_{0.025}^2(10)=20.483$. 由此可知,若 $Q\sim\chi^2(n)$,则 $P(3.247\leqslant Q\leqslant 20.483)=0.95$.

6.4.3 t 分布

定义 6.4.18 设 $X\sim N(0,1)$, $Q\sim\chi^2(n)$,且 X 与 Q 相互独立. 记 $T=X\Big/\sqrt{\frac{Q}{n}}$,则 T 的分布称为具自由度 n 的 **t 分布**,又称为**学生(student)分布**,记作 $T\sim t(n)$.

t 分布的密度函数为

$$f(x;n)=\left[\Gamma\left(\frac{n+1}{2}\right)\Big/(n\pi)^{1/2}\Gamma\left(\frac{n}{2}\right)\right]\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$$-\infty<x<+\infty.$$

$t(n)$的密度与 $N(0,1)$的密度图象比较见图 6.5.

从图 6.5 可看出,$t(n)$分布非常接近标准正态分布.

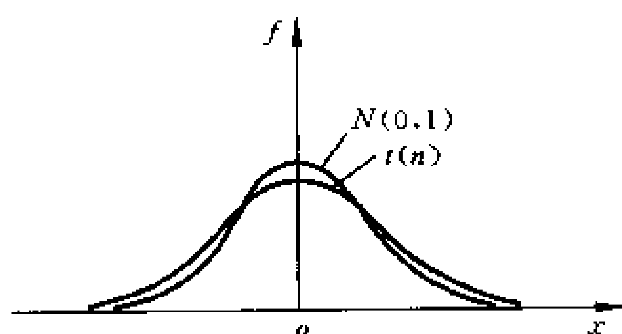


图 6.5

性质 6.4.19 $t(n)$的特征

(1) $E(T)=0$(当 $n>1$),$\sigma^2(T)=\frac{n}{n-2}$(当 $n>2$);

(2) 矩母函数不存在,特征函数无简单表达式;

(3) 0 为对称中心,中位数和众数.

定理 6.4.20 设 $X\sim N(0,1)$,$T_n\sim t(n)$,$n=1,2,\cdots$,则

$$\lim_{n\to\infty}T_n \xlongequal{d} X.$$

t 分布常用于数理统计中正态总体均值的区间估计和检验.

定理 6.4.21 设 $X_1,\cdots,X_n$ 独立,同 $N(\mu,\sigma^2)$分布,记 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$,$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$,则

$$T=\frac{\sqrt{n}\,(\overline{X}-\mu)}{S}\sim t(n-1).$$

定理 6.4.22 设 $X_1,\cdots,X_n$ 独立,同 $N(\mu_1,\sigma^2)$分布,$Y_1,\cdots,Y_m$ 独立同 $N(\mu_2,\sigma^2)$分布,且$\{X_1,\cdots,X_n\}$与$\{Y_1,\cdots,Y_m\}$独立. 定义 $\overline{X}$ 和 S_X^2 如上,定义 $\overline{Y}=\frac{1}{m}\sum_{j=1}^{m}Y_j$,$S_Y^2=\frac{1}{m-1}\sum_{j=1}^{m}(Y_j-\overline{Y})^2$,令

$$T=\frac{\sqrt{\frac{mn}{m+n}}(\overline{X}-\overline{Y})}{\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{(n+m-2)}}},$$

则当 $\mu_1=\mu_2$ 时，$T\sim t(n+m-2)$.

定义 6.4.23 设 $X\sim N(\mu,1)$，$Q\sim\chi^2(n)$，且 X 与 Q 相互独立，记 $T=X\Big/\sqrt{\dfrac{Q}{n}}$，称 T 的分布为具自由度 n 和非中心参数 μ 的非中心 t 分布，记作 $T\sim t(n,\mu)$.

$t(n)$ 为 $t(n,\mu)$ 当 $\mu=0$ 时的特例，常称为中心 t 分布.

6.4.4 F 分布

定义 6.4.24 设 $Q_1\sim\chi^2(n_1)$，$Q_2\sim\chi^2(n_2)$，且 Q_1 与 Q_2 相互独立，记 $F=\left(\dfrac{Q_1}{n_1}\right)\Big/\left(\dfrac{Q_2}{n_2}\right)$，则 F 的分布称为具自由度 (n_1,n_2) 的 **F 分布**，记作 $F\sim F(n_1,n_2)$.

$F(n_1,n_2)$ 有密度函数

$$f(x;n_1,n_2)=\begin{cases}\dfrac{(n_1/n_2)^{\frac{n_1}{2}}}{B\left(\dfrac{n_1}{2},\dfrac{n_2}{2}\right)}x^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

$F(n_1,n_2)$ 的密度图象示意如图 6.6.

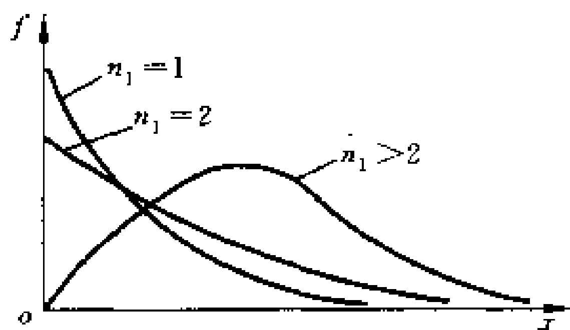


图 6.6

性质 6.4.25 $F(n_1,n_2)$ 的特征

(1) $E(F)=\dfrac{n_2}{n_2-2}$ $(n_2>2)$，

$$\sigma^2(F)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}\quad (n_2>4);$$

(2) 矩母函数不存在，特征函数无简单表达式.

F 分布在数理统计中常用于方差比的检验.

推理 6.4.26 设 S_X^2 和 S_Y^2 如定理 6.4.22 中所定义，则

$$F=\frac{S_X^2}{S_Y^2}\sim F(n-1,m-1).$$

定义 6.4.27 设 $Q_1\sim\chi^2(n_1,\delta)$，$Q_2\sim\chi^2(n_2)$，且 Q_1 与 Q_2 相互独立，记 $F=\dfrac{Q_1/n_1}{Q_2/n_2}$，则 F 的分布称为具自由度(n_1,n_2)和非中心参数 δ 的非中心 F 分布，记作 $F\sim F(n_1,n_2,\delta)$.

$F(n_1,n_2)$ 是 $F(n_1,n_2,\delta)$ 当 $\delta=0$ 时的特例，又称为中心 F 分布.

6.4.5 对数正态分布

定义 6.4.28 设 $X\sim N(\mu,\sigma^2)$，$Y=e^X$，则 Y 的分布称为**对数正态分布**(log-normal distribution).

对数正态分布有密度函数

$$f(y)=\frac{1}{\sigma y\sqrt{2\pi}}\exp\left\{-\frac{(\log y-\mu)^2}{2\sigma^2}\right\},y>0.$$

对数正态分布的密度图象示意如图 6.7.

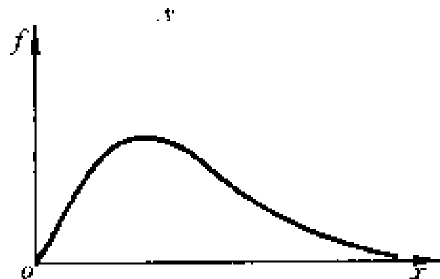


图 6.7

性质 6.4.29 对数正态分布的特征

(1) $E(Y)=\mathrm{e}^{a+\frac{\sigma^2}{2}}, \sigma^2(Y)=\mathrm{e}^{2a+\sigma^2}(\mathrm{e}^{\sigma^2}-1)$;

(2) 矩母函数不存在,特征函数无简单表达式.

对数正态分布作为寿命分布常用于可靠性统计中.

6.5 其它连续型分布

6.5.1 均匀分布

定义 6.5.1 设随机变量 X 有密度

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b,\\ 0, & \text{其它}.\end{cases}$$

则称 X 服从区间$[a,b]$上的**均匀分布**(uniform distribution),记作 $X\sim U(a,b)$.

均匀分布产生于几何概率模型,最有用的是$U(0,1)$.

性质 6.5.2 $U(0,1)$的特征

(1) $E(X)=\dfrac{1}{2}, \sigma^2(X)=\dfrac{1}{12}$;

(2) 矩母函数 $m(t)=\dfrac{\mathrm{e}^t-1}{t}$,

特征函数 $\varphi(t)=\dfrac{\mathrm{e}^{\mathrm{i}t}-1}{\mathrm{i}t}$.

定理 6.5.3 设随机变量 X 有单调连续分布函数 $F(x)$,令 $Y=F(x)$,则 $Y\sim U(0,1)$.

利用此定理可产生出服从任一指定的连续分布的随机数,假定需要产生的随机数应服从分布 F,则可先产生服从 $U(0,1)$的随机数 $y_1,y_2,\cdots,y_N$,这可以利用随机数表或计算机上的随机数发生器来作到,再令 $x_i=F^{-1}(y_i),i=1,\cdots,N$. 就得到服从分布 F 的随机数.

6.5.2 指数分布

定义 6.5.4 设随机变量 X 有概率密度

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

其中 λ 为正常数，则称 X 服从**指数分布**（exponential distribution）.

指数分布的分布函数为

$$F(x)=\begin{cases}0, & 当\ x<0,\\ 1-e^{-\lambda x}, & 当\ x\geqslant 0.\end{cases}$$

指数分布在机理上与泊松过程有关.设一块放射性物质在时间间隔 t 内放射出的 α 粒子数服从参数为 λt 的泊松分布，则 $1-e^{-\lambda t}$是在 t 时间间隔内一定会放射出 α 粒子的概率.作为 t 的函数，显然 $F(t)=1-e^{-\lambda t}$就是指数分布的分布函数.因此，指数分布是泊松过程中等待第一个非零计数出现的时间的分布.

指数分布在可靠性统计中常用作"寿命"分布，例如用来刻画无线电元件的寿命.

性质 6.5.5 指数分布的特征

(1) $E(X)=\dfrac{1}{\lambda},\sigma^2(X)=\dfrac{1}{\lambda^2}$；

(2) 矩母函数 $m(t)=\dfrac{1}{1-\left(\dfrac{t}{\lambda}\right)},t<\lambda$，

特征函数 $\varphi(t)=\dfrac{1}{1-\left(\dfrac{it}{\lambda}\right)}$.

定理 6.5.6 设 X 服从指数分布，则

$$P(X>t+s\mid X>s)=P(X>t).$$

此定理的结论常称为指数分布的"无记忆"性.假若无线电元件的寿命服从指数分布，则一个元件在已经使用了一段时间而未

损坏时，它还可以充当一个新的元件来使用.

6.5.3 Γ分布

定义 6.5.7 设随机变量 X 有密度

$$f(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

其中 λ 和 r 均为正常数，则称 X 服从 **Γ分布**，记作 $X\sim\Gamma(\lambda,r)$.

Γ 分布也是泊松过程中一个等待时间分布的推广. 当 $r=1$ 时就是指数分布. Γ 分布也在可靠性统计中用作寿命分布.

性质 6.5.8 Γ 分布的特征

(1) $E(X)=\dfrac{r}{\lambda},\sigma^2(X)=\dfrac{r}{\lambda^2}$;

(2) 矩母函数　$m(t)=\left(1-\dfrac{t}{\lambda}\right)^{-r},t<\lambda,$

特征函数　$\varphi(t)=\left(1-\dfrac{\mathrm{i}t}{\lambda}\right)^{-r}.$

Γ 分布的密度图象示意如图 6.8.

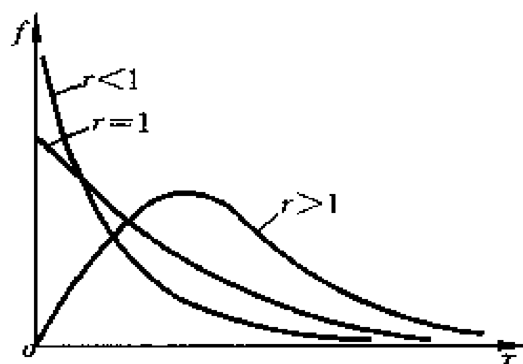


图 6.8

定理 6.5.9 设 $X_1,\cdots,X_n$ 为相互独立，具有相同参数 λ 的指数分布的随机变量，则 $\sum_{i=1}^{n}X_i$ 的分布为具参数 λ,n 的 Γ 分布.

推论 6.5.10 $\chi^2(n)$即为具参数 $\lambda=\frac{1}{2}, r=\frac{n}{2}$的 Γ 分布，特别当 $n=2$ 时，$\chi^2(2)$即为具参数 $\lambda=\frac{1}{2}$的指数分布.

6.5.4 β分布

定义 6.5.11 设随机变量 X 有密度

$$f(x)=\begin{cases}\dfrac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, & 0<x<1,\\ 0, & \text{其它}.\end{cases}$$

其中 p,q 为正常数，$B(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}\mathrm{d}x$，则称 X 服从参数为(p,q)的 β(Beta)**分布**.

性质 6.5.12 β 分布的特征

(1) $E(X)=\dfrac{p}{p+q}, \sigma^2(X)=\dfrac{pq}{(p+q)^2(1+p+q)}$;

(2) 特征函数 $\varphi(t)=\dfrac{\Gamma(p+q)}{\Gamma(p)}\sum\limits_{k=0}^{\infty}\dfrac{\Gamma(p+q)(\mathrm{i}t)^k}{\Gamma(p+q+k)\Gamma(k+1)}$.

定理 6.5.13 设 $X_1\sim\Gamma(\lambda,p), X_2\sim\Gamma(\lambda,q)$(见定义 6.5.7)，且 X_1 与 X_2 相互独立，则$\dfrac{X_1}{X_1+X_2}$服从参数为(pq)的 β 分布.

推论 6.5.14 设 $X\sim F(n,m)$，则$(1+X)^{-1}$服从参数为$\left(\dfrac{m}{2},\dfrac{n}{2}\right)$的 β 分布.

β 分布在数理统计的多元分析中有重要应用.

6.5.5 韦布尔分布

定义 6.5.15 设随机变量 X 有密度

$$f(x)=\begin{cases}\dfrac{r}{\alpha}\left(\dfrac{x-\mu}{\alpha}\right)^{r-1}\mathrm{e}^{-\left(\frac{x-\mu}{\alpha}\right)^r} & x\geqslant\mu,\\ 0 & x<\mu.\end{cases}$$

其中 α 和 r 为正常数，μ 为实常数，则称 X 服从**韦布尔**(weibull)**分布**，记作 $X \sim W(\alpha, r, \mu)$.

推论 6.5.16 设 $X \sim W(\alpha, r, \mu)$，则 $\left(\dfrac{X-\mu}{\alpha}\right)^r$ 服从参数 $\lambda = 1$ 的指数分布.

韦布尔分布的密度图象示意如图 6.9.

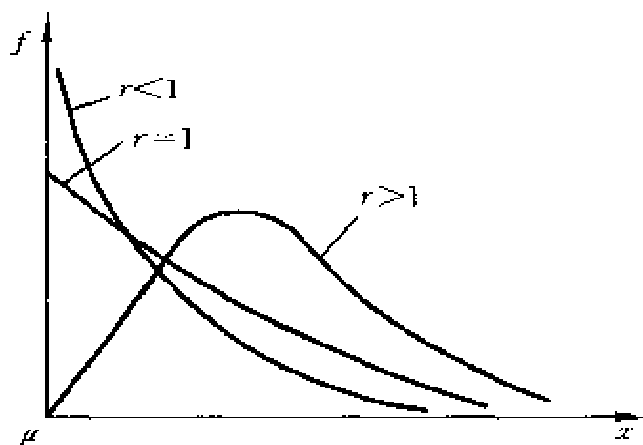


图 6.9

性质 6.5.17 $W(\alpha, r, \mu)$的特征

(1) $E(X) = \alpha\Gamma\left(1+\dfrac{1}{r}\right)+\mu$,

$$\sigma^2(X) = \alpha^2\left[\Gamma\left(1+\frac{2}{r}\right) - \Gamma^2\left(1+\frac{1}{r}\right)\right];$$

(2) 矩母函数与特征函数无简单形式.

韦布尔分布是可靠性理论中的基本分布.

6.5.6 瑞利分布

定义 6.5.18 设随机变量 X 有密度

$$f(x) = \begin{cases} \dfrac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

其中 $\sigma > 0$，则称 X 服从**瑞利**(Rayleigh)**分布**.

定理 6.5.19 设 X_1, X_2 为相互独立，具相同分布 $N(0,\sigma^2)$ 的随机变量，则 $\sqrt{X_1^2+X_2^2}$ 的分布为瑞利分布.

性质 6.5.20 瑞利分布的特征

(1) $E(X)=\sqrt{\frac{\pi}{2}}\sigma, \sigma^2(X)=\frac{4-\pi}{2}\sigma^2$;

(2) 矩母函数与特征函数无简单形式.

推论 6.5.21 瑞利分布为 $W(2,\sqrt{2}\sigma,0)$：韦布尔分布的一个特例.

6.5.7 拉普拉斯分布

定义 6.5.22 设随机变量 X 有密度函数

$$f(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}, \quad -\infty < x < \infty.$$

其中 λ 为正常数，μ 为实常数，则称 X 服从**拉普拉斯**(Laplace)**分布**.

性质 6.5.23 拉普拉斯分布的特征

(1) $E(X)=\mu, \sigma^2(X)=2\lambda^2$;

(2) 特征函数 $\varphi(t)=\frac{e^{i\mu t}}{1+\lambda^2t^2}$.

6.5.8 柯西分布

定义 6.5.24 设随机变量 X 有密度

$$f(x)=\frac{\lambda}{\pi[\lambda^2+(x-\mu^2)]}, \quad -\infty < x < \infty,$$

其中 λ 为正常数，μ 为实常数，则称 X 服从**柯西**(Cauchy)**分布**.

定理 6.5.25 设 θ 为 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 上均匀分布随机变量，则 $X=\operatorname{tg}\theta$ 的分布为具参数 $\lambda=1,\mu=0$ 的柯西分布.

性质 6.5.26 柯西分布的特征

(1) 各阶矩都不存在；

(2) 特征函数 $\varphi(t)=e^{i\mu t-\lambda|t|}$.

柯西分布的密度图象示意如图 6.10.

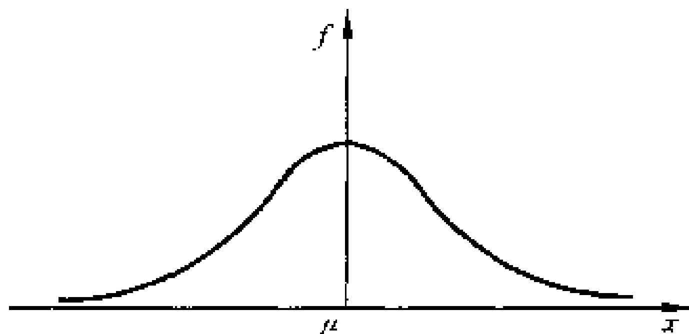


图 6.10

定理 6.5.27 设 $X_1,X_2,\cdots,X_n$ 为相互独立，有相同柯西分布的随机变量，则 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 也服从柯西分布.

由此定理可见，中心极限定理对柯西分布的随机变量列不成立.

6.5.9 帕雷托分布

定义 6.5.28 设随机变量 X 有密度

$$f(x)=\begin{cases}\dfrac{r}{(1+x)^{r+1}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

其中 r 为正常数，则称 X 服从**帕雷托**(Pareto)**分布**.

性质 6.5.29 帕雷托分布的特征

(1) $E(X)=\frac{1}{r-1}\quad(r>1)$，$\sigma^2(X)=\frac{r}{(r-1)(r-2)}\quad(r>2)$；

(2) 矩母函数不存在，特征函数无简单形式.

帕雷托分布常用于经济学中，帕雷托首先用它描述财政收入的分布.

6.5.10 逻辑斯谛分布

定义 6.5.30 设随机变量 X 有密度

$$f(x)=\frac{e^{-x}}{(1+e^{-x})^2},\ -\infty<x<\infty,$$

则称 X 服从**逻辑斯谛**(Logistic)**分布**.

性质 6.5.31 逻辑斯谛分布的特征

(1) $E(X)=0,\sigma^2(X)=\frac{\pi^2}{3}$;

(2) 矩母函数 $m(t)=\pi t\operatorname{cosec}(\pi t),-\infty<t<\infty$,

特征函数 $\varphi(t)=\pi it\operatorname{cosec}(\pi it)$.

逻辑斯谛分布常用于经济学与人口统计学.

6.6 多元分布

在多元分布中,最重要的是多元正态分布,在数理统计的多元分析领域中有广泛的应用.这部分内容将在第 18 章 18.1 中详细介绍.

6.6.1 二元正态分布

定义 6.6.1 设二维随机向量(X,Y)有联合密度

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},$$

$$-\infty<x,y<+\infty.$$

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 均为常数,分别满足 $-\infty<\mu_1,\mu_2<\infty,\sigma_1>0,\sigma_2>0,|\rho|<1$,则称随机向量$(X,Y)$的分布为**二元正态分布**,记作 $(X,Y)\sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

性质 6.6.2 二元正态分布的特征

(1) $E(X)=\mu_1$, $E(Y)=\mu_2$,

$\sigma^2(X)=\sigma_1^2$, $\sigma^2(Y)=\sigma_2^2$, $\rho(X,Y)=\rho$;

(2) 特征函数

$$\varphi(t_1,t_2)=\exp\left\{\mathrm{i}(\mu_1 t_1+\mu_2 t_2)-\frac{1}{2}\left[t_1^2\sigma_1^2+2t_1t_2\rho\sigma_1\sigma_2+t_2^2\sigma_2^2\right]\right\}.$$

二元正态密度图象示意如图 6.11.

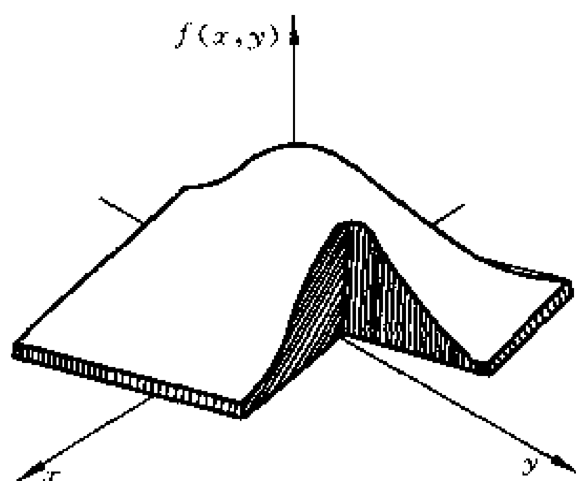


图 6.11

二元正态分布的边缘分布与条件分布仍是正态分布.

$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$,

$$X|Y=y\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2)\right),$$

$$Y|X=x\sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\right).$$

定理 6.6.3 设(X,Y)服从二元正态分布,则 X 与 Y 独立的充分必要条件为 X 与 Y 不相关,即 $\rho=0$.

6.6.2 多项分布

定义 6.6.4 设随机向量$(X_1,X_2,\cdots,X_r)$有概率分布

$$P(X_1=k_1,\cdots,X_r=k_r)=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r},$$

其中 $k_1,\cdots,k_r$ 为非负整数，满足 $\sum_{i=1}^{r}k_r=n$，$p_1,\cdots,p_r$ 为正常数，满足 $\sum_{i=1}^{r}p_i=1$，则称此分布为**多项分布**(multinomial distribution).

多项分布可由一个 n 重试验产生，其中每次试验产生 r 个基本事件 $A_1,\cdots,A_r$. 记 $p_i=P\{A_i\}$，$i=1,\cdots,r$，$\sum_{i=1}^{r}p_i=1$，这是 n 重伯努利试验的推广. 记 X_i 为 A_i 在 n 次试验中发生的次数，$i=1,\cdots,r$，则 $(X_1,X_2,\cdots,X_r)$ 服从多项分布.

由定义可知，多项分布中实际上只有 $r-1$ 个自由变量. 即

$$P\{X_1+X_2+\cdots+X_r=n\}=1.$$

性质 6.6.5 多项分布的特征

(1) $E(X_i)=np_i$，$i=1,\cdots,r$；

$\sigma^2(X_i)=np_i(1-p_i)$，$i=1,\cdots,r$；

$\mathrm{cov}(X_i,X_j)=-np_ip_j$，$i,j=1,\cdots,r$，$i\neq j$；

(2) 特征函数

$$\varphi(t_1,\cdots,t_r)=(p_1\mathrm{e}^{\mathrm{i}t_1}+\cdots+p_r\mathrm{e}^{\mathrm{i}t_r})^n.$$

定义 6.6.6 边缘分布与条件分布

(1) 对 $s<r$，$(X_1,\cdots,X_s)$ 的边缘分布为

$$P(X_1=k_1,\cdots,X_s=k_s)$$

$$=\frac{n!}{k_1!\cdots k_s!\left(n-\sum_{1}^{s}k_i\right)!}p_1^{k_1}\cdots p_s^{k_s}\left(1-\sum_{1}^{s}p_i\right)^{n-\sum_{1}^{s}k_i},$$

其中 $k_1+\cdots+k_s\leqslant n$，特别有 $X_i\sim B(n,p_i)$，$i=1,\cdots,r$；

(2) 给定 $X_{s+1}=k_{s+1},\cdots,X_r=k_r$ 时，$X_1,\cdots,X_s$ 的条件分布为

$$P(X_1=k_1,\cdots,X_s=k_s\mid X_{s+1}=k_{s+1},\cdots,X_r=k_r)$$

$$= \frac{\left(n - \sum_{i=s+1}^{r} k_i\right)!}{k_1! \cdots k_s!} p_1^{k_1} \cdots p_s^{k_s} \left(\sum_{i=1}^{s} p_i\right)^{-\sum_{i=1}^{s} k_i},$$

其中 $k_1 + \cdots + k_s = n - \sum_{i=s+1}^{r} k_i$.

例 6.6.7 设一批产品可分为 4 类:一等品,二等品,三等品和不合格品,各占比例 $100p_i\%$,$i=1,2,3,4$, $\sum_{i=1}^{4} p_i = 1$. 从这批产品中有放回地抽取 n 个,用 $X_1,\cdots,X_4$ 记样品中一、二、三等品及不合格品的个数,则 $X_1,\cdots,X_4$ 的分布为 4 项分布.

6.6.3 狄利克雷分布

定义 6.6.8 设随机向量 $(X_1, X_2, \cdots, X_n)$ 满足 $P\left(\sum_{1}^{n} X_i = 1\right) = 1$,且 $(X_1,X_2,\cdots,X_{n-1})$ 有概率密度函数

$$f(x_1,\cdots,x_{n-1}) = \begin{cases} \dfrac{\Gamma\left(\sum_{1}^{n} r_i\right)}{\prod_{1}^{n} \Gamma(r_i)} \prod_{1}^{n-1} x_i^{r_i-1}\left(1 - \sum_{1}^{n-1} x_i\right)^{r_n-1} & x_i > 0, i = 1,\cdots,n-1, \\ & \sum_{1}^{n-1} x_i \leqslant 1, \\ 0 & \text{其它}. \end{cases}$$

其中 r_i 为正常数. $i=1,\cdots,n$,则 $(X_1,\cdots,X_n)$ 的分布称为**狄利克雷(Dirichlet)分布**,记作 $(X_1,\cdots,X_n) \sim D_n(r_1,\cdots,r_n)$.

在 $(0,1)$ 区间上随机地取 $n-1$ 个点 $P_1,P_2,\cdots,P_{n-1}$,将此 $n-1$ 个点依从小到大的顺序记为 $P_{(1)},P_{(2)},\cdots,P_{(n-1)}$. 记 $X_i = P_{(i)} - P_{(i-1)}$,$i=1,\cdots,n$;其中 $P_{(0)}=0$,$P_{(n)}=1$,则 $(X_1,\cdots,X_n) \sim D_n(1,1,\cdots,1)$.

性质 6.6.9 $D_n(r_1,\cdots,r_n)$ 的特征

(1) $E(X_i)=\dfrac{r_i}{\sum\limits_{j=1}^{n} r_j}, \sigma^2(X_i)=\dfrac{r_i \sum\limits_{j\neq i} r_j}{\left(\sum\limits_{j=1}^{n} r_j\right)^2\left(1+\sum\limits_{j=1}^{n} r_j\right)},$

$$i=1,\cdots,n;$$

(2) $\operatorname{cov}(X_i,X_j)=\dfrac{-r_i r_j}{\left(\sum\limits_{k=1}^{n} r_k\right)^2\left(\sum\limits_{k=1}^{n} r_k+1\right)},\quad i\neq j.$

定理 6.6.10 设 $X_1,\cdots,X_n$ 相互独立, $X_i\sim\chi^2(k_i), i=1,\cdots,n$. 令

$$Y_i=X_i\Big/\sum_{j=1}^{n} X_j, i=1,\cdots,n,$$

则

$$(Y_1,\cdots,Y_n)\sim D_n\left(\frac{k_1}{2},\cdots,\frac{k_n}{2}\right).$$

7 概率测度与积分

在本书的前几章中，对随机事件，概率及随机变量等一些重要的基本概念作了较为直观的介绍，目的是使实际工作者易于理解掌握，不足之处是数学上不够严格. 本章对概率论的公理化体系的若干基本要点作一概括介绍，它是前苏联著名数学家科尔莫戈罗夫(Колмогоров)于 20 世纪 30 年代建立起来的.

7.1 测度空间与概率空间

一个由研究对象全体构成的集合称为**样本空间**，简称**空间**，记作 Ω. Ω 中的元素记作 ω.

定义 7.1.1 空间 Ω 上的一个非空子集类 $\mathscr{F}$ 若满足条件：

(1) $A\in\mathscr{F}\Rightarrow\overline{A}\in\mathscr{F}$,

(2) $A_1\in\mathscr{F},A_2\in\mathscr{F}\Rightarrow A_1\cup A_2\in\mathscr{F}$,

则称 $\mathscr{F}$ 为一个**域**(field)，又称为**代数**(algebra).

推论 7.1.2 设 $\mathscr{F}$ 为空间 Ω 上的一个域，则

(1) $\Omega\in\mathscr{F},\varnothing\in\mathscr{F}$;

(2) $A_j\in\mathscr{F},j=1,2,\cdots,n\Rightarrow \bigcup\limits_{j=1}^{n}A_j\in\mathscr{F}$, $\bigcap\limits_{j=1}^{n}A_j\in\mathscr{F}$.

定义 7.1.3 设 $\mathscr{F}$ 为 Ω 上的一个域，且满足

$$A_j\in\mathscr{F},j=1,2,\cdots\Rightarrow \bigcup_{j=1}^{\infty}A_j\in\mathscr{F},$$

则称 $\mathscr{F}$ 为一个 **σ 域**，或 **σ 代数**.

推论 7.1.4 若 $\mathscr{F}$ 为一个 σ-域，则

$$A_j\in\mathscr{F},j=1,2,\cdots\Rightarrow \bigcap_{j=1}^{\infty}A_j\in\mathscr{F}.$$

定义 7.1.5 空间 Ω 与其上的一个 σ 域 $\mathscr{F}$ 构成一个**可测空间**(measurable space)，记作 $(\Omega,\mathscr{F})$，$\mathscr{F}$ 中的任一元素(Ω 中的子集)称为**可测集**(measurable set).

定义 7.1.6 设 $(\Omega,\mathscr{F})$ 为可测空间，μ 为定义在 $\mathscr{F}$ 上的非负(集合)函数，满足：

(1) $\mu(\varphi)=0$,

(2) 若 $A_n\in\mathscr{F}$，$n=1,2,\cdots$，且 $A_nA_m=\varnothing$，当 $n\neq m$，则 $\mu\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\mu(A_n)$.

称 μ 为 $\mathscr{F}$ 上的**测度**(measure)，称 $(\Omega,\mathscr{F},\mu)$ 为**测度空间**(measure space).

定义 7.1.7 设 P 为可测空间 $(\Omega,\mathscr{F})$ 上的测度，且 $P(\Omega)=1$，则称 P 为 $(\Omega,\mathscr{F})$ 上的**概率测度**，$(\Omega,\mathscr{F},P)$ 三位一体称为**概率空间**.

定义 7.1.6 中性质(2)称为测度的 **σ 可加性**.

定义一个测度空间(或概率空间)的基本方法是先对一个较小的集合类定义其上的测度，然后进行扩张.

定义 7.1.8 对空间 Ω 中的一个非空子集类 $\mathscr{A}$，若 $\mathscr{F}$ 为一个 σ 域，满足：

(1) $\mathscr{F}\supset\mathscr{A}$,

(2) 对任一包含 $\mathscr{A}$ 的 σ 域 $\mathscr{F}'$，有 $\mathscr{F}'\supset\mathscr{F}$，则称 $\mathscr{F}$ 为**由 A 生成的 σ 域**(σ field generated from $\mathscr{A}$)，记作 $\mathscr{F}=\sigma(\mathscr{A})$.

$\sigma(\mathscr{A})$ 又称为包含 $\mathscr{A}$ 的最小 σ 域.

定义 7.1.9 设 $\mathscr{A}$ 为实数空间 $\mathbf{R}^1$ 上全体左开右闭区间构成的子集类，$\mathscr{B}^1=\sigma(\mathscr{A})$：由 $\mathscr{A}$ 生成的 σ 域，则 $(\mathbf{R}^1,\mathscr{B}^1)$ 称为**博雷尔**(Borel)**空间**，$\mathscr{B}$ 中的元素($\mathbf{R}^1$ 中的子集)称为**博雷尔集**.

定义 7.1.10 设 $\mathscr{F}$ 和 $\mathscr{A}$ 为空间 Ω 上的两个子集类，$\mathscr{F}\supset\mathscr{A}$，$\nu$ 为 $\mathscr{A}$ 上的测度，μ 为 $\mathscr{F}$ 上的测度，若对 $\forall\ A\in\mathscr{A}$ 有 $\mu(A)$

$=\nu(A)$，则称 μ 为 ν 由 $\mathscr{A}$ 到 $\mathscr{F}$ 上的一个**扩张**(extension).

定义 7.1.11 设 $\mathscr{A}$ 为空间 Ω 上的子集类，μ 为 $\mathscr{A}$ 上的测度，且存在 $A_n \in \mathscr{A}, n=1,2,\cdots$, $\bigcup_{n=1}^{\infty} A_n = \Omega$, $\mu(A_n)<\infty, n=1,2,\cdots$，则称 μ 为 **σ 有限测度**(σ finit measure).

定理 7.1.12 设

(1) $\mathscr{A}$ 为空间 Ω 上的域，

(2) μ 为 $\mathscr{A}$ 上的 σ 有限测度，

则存在 μ 由 $\mathscr{A}$ 到 $\sigma(\mathscr{A})$上的唯一的扩张.

定理 7.1.12 是定义一些常用的测度和概率测度的依据. 例如，设 $f(x)$为 $\mathbf{R}^1$ 上非负可积函数，对任一左开右闭区间$(a,b]$，定义 $\mu((a,b]) = \int_a^b f(x)\mathrm{d}x$ ，则 μ 在博雷尔空间$(\mathbf{R}^1,\mathscr{B}^1)$上的扩张就是一个测度，且当 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ 时，就是一个概率测度.

定义 7.1.13 设$(\Omega_i,\mathscr{F}_i)$为可测空间，$i=1,2,\cdots,n$，定义

(1) $\Omega_1\times\Omega_2\times\cdots\times\Omega_n=\{(\omega_1,\omega_2,\cdots,\omega_n);\omega_i\in\Omega_i,1\leqslant i\leqslant n\}$，称为 n 维**乘积空间**(product space)，简记为 $\bigtimes_{i=1}^{n}\Omega_i$；

(2) 对 $A_i\in\mathscr{F}_i,1\leqslant i\leqslant n$，定义 $A_1\times A_2\times\cdots\times A_n=\{(\omega_1,\omega_2,\cdots,\omega_n),\omega_i\in A_i,1\leqslant i\leqslant n\}$，称为 **$n$ 维矩形**，简记为 $\bigtimes_{i=1}^{n}A_i$；

(3) 对由所有 n 维矩形构成的子集类 $\mathscr{A}$，定义 $\mathscr{F}_1\times\mathscr{F}_2\times\cdots\times\mathscr{F}_n$ 为由 $\mathscr{A}$ 生成的 σ 域，称为 **n 维乘积 σ 域**，简记为 $\bigtimes_{i=1}^{n}\mathscr{F}_i$；

(4) 由 n 维乘积空间 $\bigtimes_{i=1}^{n}\Omega_i$ 及其上的 n 维乘积 σ 域 $\bigtimes_{i=1}^{n}\mathscr{F}_i$ 构成的可测空间 $(\bigtimes_{i=1}^{n}\Omega_i, \bigtimes_{i=1}^{n}\mathscr{F}_i)$ 称为 **n 维可测空间**，记作 $\bigtimes_{i=1}^{n}(\Omega_i,\mathscr{F}_i)$.

定义 7.1.14 设$\langle \Omega_n, \mathscr{F}_n \rangle$为可测空间，$n=1,2,\cdots$，定义

(1) $\bigtimes_{n=1}^{\infty} \Omega_n = \{(\omega_1, \omega_2, \cdots), \omega_n \in \Omega_n, n=1,2,\cdots\}$，称为**无穷维乘积空间**.

(2) 对 $A_i \in \mathscr{F}_i, i=1,\cdots,n$，定义

$$\left(\bigtimes_{i=1}^{n} A_i\right) \times \left(\bigtimes_{i=n+1}^{\infty} \Omega_i\right)$$
$$=\{(\omega_1,\cdots,\omega_n,\omega_{n+1},\cdots), \omega_i \in A_i, i=1,\cdots,n,$$
$$\omega_j \in \Omega_j, j=n+1, n+2, \cdots\},$$

称为 n 维**柱形集**(cylinder).

(3) 记 $\mathscr{A}$ 为集合类$\{A \mid A$ 为 n 维柱形集，$n=1,2,\cdots\}$，定义 $\bigtimes_{n=1}^{\infty} \mathscr{F}_n$ 为由 $\mathscr{A}$ 生成的 σ 域 $\sigma(\mathscr{A})$，称为**无穷维 σ 域**.

(4) 由无穷维可测空间 $\bigtimes_{n=1}^{\infty} \Omega_n$ 及其上的无穷维乘积 σ 域 $\bigtimes_{n=1}^{\infty} \mathscr{F}_n$ 构成的可测空间$\left(\bigtimes_{n=1}^{\infty} \Omega_n, \bigtimes_{n=1}^{\infty} \mathscr{F}_n\right)$称为**无穷维可测空间**，记作 $\bigtimes_{n=1}^{\infty} (\Omega_n, \mathscr{F}_n)$.

定义 7.1.15 设$\langle \mathrm{R}^1, \mathscr{B}^1 \rangle$为博雷尔空间，定义

(1) **n 维博雷尔空间**$\langle R^n, \mathscr{B}^n \rangle = \underbrace{\langle R, \mathscr{B} \rangle \times \cdots \times \langle R, \mathscr{B} \rangle}_{n\text{个}}$

(2) **无穷维博雷尔空间**$\langle R^\infty, \mathscr{B}^\infty \rangle = \langle R, \mathscr{B} \rangle \times \langle R, \mathscr{B} \rangle \times \cdots\cdots$.

对应第 1 章中的内容，一个有限(或无穷)维可测空间表示一个有限(或无穷)多个随机试验序列的样本空间.

定理 7.1.16 设$\langle \Omega_i, \mathscr{F}_i, \mu_i \rangle i=1,2,\cdots$为有限或可列无穷多个 σ 有限测度空间，则在乘积可测空间$\langle \bigtimes_i \Omega_i, \bigtimes_i \mathscr{F}_i \rangle$上存在唯

一的 σ 有限测度，记作 $\underset{i}{\times}\mu_i$，满足

$$\underset{i}{\times}\mu_i(\overset{n}{\underset{i=1}{\times}}A_i)=\prod_{i=1}^{n}\mu_i(A_i),A_i\in\mathscr{F}_i,1\leqslant i\leqslant n,n=1,2,\cdots,$$

称 $\underset{i}{\times}\mu_i$ 为乘积测度，$\langle\underset{i}{\times}\Omega_i,\ \underset{i}{\times}\mathscr{F}_i,\ \underset{i}{\times}\mu_i\rangle$ 为 **乘积测度空间**.

定义 7.1.17 设 $\langle\Omega_i,\mathscr{F}_i,P_i\rangle,i=1,2,\cdots$ 为有限或可列无穷多个概率空间，$\langle\underset{i}{\times}\Omega_i,\ \underset{i}{\times}\mathscr{F}_i,\ \underset{i}{\times}P_i\rangle$ 称为**乘积概率空间**.

乘积概率空间对应着独立的随机试验序列，即对任给正整数 n，及任意 n 个事件 $A_i\in\mathscr{F}_i,i=1,\cdots,n$，有

$$P(A_1A_2\cdots A_n)=P_1(A_1)\cdots P_n(A_n),$$

其中 P 表示乘积概率 $\overset{n}{\underset{i=1}{\times}}P_i$.

7.2 可测函数与随机变量

定义 7.2.1 设 X 为空间 Ω_1 到空间 Ω_2 的映射，对 Ω_2 的任一子集 A，定义 Ω_1 中的子集 $X^{-1}(A)=\{\omega_1:\omega_1\in\Omega_1,X(\omega_1)\in A\}$，$X^{-1}(A)$称为 **$A$ 的逆象**(inverse image). 设 $\mathscr{A}$ 为 Ω_2 的一个子集类，定义 Ω_2 中相应的子集类 $X^{-1}(\mathscr{A})=\{X^{-1}(A):A\in\mathscr{A}\}$，$X^{-1}(\mathscr{A})$称为 **$\mathscr{A}$ 的逆象**.

定理 7.2.2 设 X 为空间 Ω_1 到空间 Ω_2 的映射.

(1) 若 $\mathscr{F}_2$ 为 Ω_2 上的 σ 域，则 $X^{-1}(\mathscr{F}_2)$为 Ω_1 上的 σ 域；

(2) 若 $\mathscr{F}_1$ 为 Ω_1 上的 σ 域，则 $\mathscr{F}_2=\{A:A\in\Omega_2,X^{-1}(A)\in\mathscr{F}_1\}$为 Ω_2 上的 σ 域；

(3) 对 Ω_2 上的任一子集类 $\mathscr{A}$，$X^{-1}(\sigma(\mathscr{A}))=\sigma(X^{-1}(\mathscr{A}))$.

定义 7.2.3 设$\langle\Omega_1,\mathscr{F}_1\rangle$，$\langle\Omega_2,\mathscr{F}_2\rangle$为两个可测空间，$X$ 为 Ω_1 到 Ω_2 的映射. 若 $X^{-1}(\mathscr{F}_2)\subset\mathscr{F}_1$，则称 X 为**可测变换**(measurable transformation).

定义 7.2.4 设 X 为从可测空间 $\langle\Omega,\mathscr{F}\rangle$ 到 1 维博雷尔空间 $\langle R^1,\mathscr{B}^1\rangle$ 的可测变换，则 X 称为**可测函数**(measurable function)；当 $\langle\Omega,\mathscr{F}\rangle$ 为 n 维博雷尔空间 $\langle R^n,\mathscr{B}^n\rangle$ 时，称 X 为 **n 维博雷尔可测函数**，简称为 **n 维博雷尔函数**.

定义 7.2.5 设 $\langle\Omega,\mathscr{F},P\rangle$ 为一概率空间，X 为 $\langle\Omega,\mathscr{F},P\rangle$ 到 n 维博雷尔空间 $\langle \mathbf{R}^n,\mathscr{B}^n\rangle$ 的可测变换，$n=1,2,\cdots,\infty$，当 $n=1$ 时称 X 为**随机变量**(random variable)，当 $n>1$ 时，称 X 为 **n 维随机向量**(n-dimensinal random vector).

定义 7.2.6 对空间 Ω 上的子集 A，定义函数

$$I_A(\omega)=\begin{cases}1, & \text{当 } \omega\in A,\\ 0, & \text{当 } \omega\overline{\in} A.\end{cases}$$

称 I_A 为 A 的**示性函数**(indicator function).

定义 7.2.7 设可测空间 $\langle\Omega,\mathscr{F}\rangle$ 上的实函数 X 可表为 $X=\sum_{i=1}^{m}c_iI_{A_i}$，其中 m 为正整数，$c_1,\cdots,c_m$ 为实常数，$A_i\in\mathscr{F}$，$i=1,\cdots,m$，且 $\bigcup_{i=1}^{m}A_i=\Omega$，$I_{A_i}$ 为 A_i 的示性函数，$i=1,\cdots,m$，则 X 称为**简单函数**(simple function).

推论 7.2.8 简单函数是可测函数.

定理 7.2.9 X 为非负可测函数的充分必要条件为，存在单调非降的简单函数列 X_n，$n=1,2,\cdots$，使得 $\lim\limits_{n\to\infty}X_n=X$.

推论 7.2.10 设 c 为常数，$X,X_1,X_2,\cdots$ 均为可测空间 $\langle\Omega,\mathscr{F}\rangle$ 上的可测函数，则

(1) cX，$|X|$，X_1+X_2，X_1X_2，X_1/X_2 均为可测函数；

(2) $X^+=\max(0,X)$ 与 $X^-=\max(0,-X)$ 均为可测函数；

(3) $\limsup\limits_{n\to\infty}(X_n)$ 与 $\liminf\limits_{n\to\infty}(X_n)$ 均为可测函数，故若 $\lim\limits_{n\to\infty}X_n$ 存在，则 $\lim\limits_{n\to\infty}X_n$ 亦为可测函数；

(4) 设 $\langle\Omega_1,\mathscr{F}_1\rangle$，$\langle\Omega_2,\mathscr{F}_2\rangle$ 和 $\langle\Omega_3,\mathscr{F}_3\rangle$ 为三个可测空间，X 为

$\langle \Omega_1, \mathscr{F}_1 \rangle$到$\langle \Omega_2, \mathscr{F}_2 \rangle$的可测变换，$g$为$\langle \Omega_2, \mathscr{F}_2 \rangle$到$\langle \Omega_3, \mathscr{F}_3 \rangle$的可测变换，则$g(x)$为$\langle \Omega_1, \mathscr{F}_1 \rangle$到$\langle \Omega_3, \mathscr{F}_3 \rangle$的可测变换.

定理 7.2.11 X为从可测空间$\langle \Omega, \mathscr{F} \rangle$到$n$维博雷尔空间$\langle \mathbf{R}^n, \mathscr{B}^n \rangle$上的可测变换，当且仅当$X$的$n$个分量分别为$\langle \Omega, \mathscr{F} \rangle$上的可测函数.

推论 7.2.12 $X=(X_1, X_2, \cdots, X_n)$为$n$维随机向量，当且仅当$X_1, \cdots, X_n$为随机变量.

定理 7.2.13 设X为由测度空间$\langle \Omega_1, \mathscr{F}_1, \mu_1 \rangle$到可测空间$\langle \Omega_2, \mathscr{F}_2 \rangle$上的可测变换，则$X$在$\langle \Omega_2, \mathscr{F}_2 \rangle$上导出一个测度$\mu_2$，满足

$$\mu_2(B) = \mu_1(X^{-1}(B)), \forall B \in \mathscr{F}_2.$$

定义 7.2.14 设X为概率空间$\langle \Omega, \mathscr{F}, P \rangle$上的$n$维随机向量，$n=1,2,\cdots,\infty$，则$X$在$\langle \mathbf{R}^n, \mathscr{B}^n \rangle$上导出一个概率测度$P$，称为$X$的**概率分布**或简称为**分布**.

定义 7.2.15 $\mathbf{R}^n$上满足下述条件的实函数F称为**n元分布函数**：

(1) $\lim\limits_{x_j \to -\infty} F(x_1, \cdots, x_n) = 0, j = 1, \cdots, n$;

(2) $\lim\limits_{\substack{x_j \to \infty \\ j=1,\cdots,n}} F(x_1, \cdots, x_n) = 1$;

(3) $\lim\limits_{y_j \to x_j^+} F(x_1, \cdots, x_{j-1}, y_j, x_{j+1}, \cdots, x_n) = F(x_1, \cdots, x_n)$,

$$j=1, \cdots, n;$$

(4) 对任意的$a_i \leqslant b_i, i=1, \cdots, n$,

$$F(b_1, \cdots, b_n) - \sum_i F(b_1, \cdots, b_{i-1}, a_i, b_{i+1}, \cdots, b_n)$$

$$+ \sum_{i<j} F(b_1, \cdots, b_{i-1}, a_i, b_{i+1}, \cdots, b_{j-1}, a_j, b_{j+1}, \cdots, b_n) - \cdots$$

$$+ (-1)^n F(a_1, \cdots, a_n) \geqslant 0.$$

定理 7.2.16 任一n元分布函数F在n维博雷尔空间

$\langle R^n,\mathscr{B}^n\rangle$上定义了一个概率测度 P,满足如下关系:

$$
\begin{aligned}
&P(\mathop{\times}_{i=1}^{n}(a_i,b_i])\\
=&F(b_1,\cdots,b_n)-\sum_i F(b_1,\cdots,b_{i-1},a_i,b_{i+1},\cdots,b_n)\\
&+\sum_{i<j}F(b_1,\cdots,b_{i-1},a_i,b_{i+1},\cdots,b_{j-1},a_j,b_{j+1},\cdots,b_n)-\cdots\\
&+(-1)^nF(a_1,\cdots,a_n),\forall\, a_i\leqslant b_i,i=1,\cdots,n.
\end{aligned}
$$

反之亦真.

定义 7.2.17 设$\{F_n|n\geqslant 1\}$为一分布函数族,其中 F_n 为 n 元分布函数,$n=1,2,\cdots$,且满足

$$\lim_{x_{n+1}\to\infty}F_{n+1}(x_1,\cdots,x_n,x_{n+1})=F_n(x_1,\cdots,x_n),$$

$\forall\, n,x_1,\cdots,x_n$. 称$\{F_n,n\geqslant 1\}$为**相容分布函数族**(family of consistent distribution function).

定理 7.2.18 科尔莫戈罗夫(Колмогоров)相容性定理 设$\{F_n,n\geqslant 1\}$为一相容分布函数族,则$\{F_n|n\geqslant 1\}$在无穷维博雷尔空间$\langle \mathbf{R}^\infty,\mathscr{B}^\infty\rangle$上唯一确定了一个概率测度 P,使得

$$P\left(\mathop{\times}_{i=1}^{n}(-\infty,x_i]\times(\mathop{\times}_{i=n+1}^{\infty}R^1)\right)=F_n(x_1,\cdots,x_n),$$
$$\forall\, n,x_1,\cdots,x_n.$$

7.3 积分

定义 7.3.1 设$\langle\Omega,\mathscr{F},\mu\rangle$为测度空间,$X$ 为$\langle\Omega,\mathscr{F}\rangle$上的可测函数,则 X 关于 μ 的**积分** $\int X\mathrm{d}\mu$ 规定如下:

(1) 若 X 为非负简单函数,$X=\sum_{i=1}^{m}c_iI_{A_i}$,其中 $A_1,\cdots,A_m$ 两两互不相交,则

$$\int X \mathrm{d}\mu = \sum_{i=1}^{m} c_i \mu(A_i);$$

(2) 设 X 为非负可测函数，若$\{X_n, n \geqslant 1\}$为单调上升的非负简单函数，满足$\lim\limits_{n\to\infty} X_n = X$，则

$$\int X \mathrm{d}\mu = \lim_{n\to\infty} \int X_n \mathrm{d}\mu;$$

(3) 对一般可测函数，若$\int X^+ \mathrm{d}\mu < \infty$或$\int X^- \mathrm{d}\mu < \infty$，则

$$\int X \mathrm{d}\mu = \int X^+ \mathrm{d}\mu - \int X^- \mathrm{d}\mu.$$

注 在上述定义中，(1)中非负简单函数 X 的表示方式不是唯一的，但积分定义式右端的值是唯一确定的；(2)中积分定义式右端的极限值对满足条件的不同的简单函数列是唯一确定的；对不满足条件(1)～(3)的可测函数 X，积分$\int X \mathrm{d}\mu$不存在.

定义 7.3.2 若$\left|\int X \mathrm{d}\mu\right| < \infty$，则称 X 为 μ **可积函数**(μ-integrable function).

推论 7.3.3 设 X, Y 为$\langle \Omega, \mathscr{F}, \mu \rangle$上可测函数，则有

(1) $\mu(X < 0) = 0 \Rightarrow 0 \leqslant \int X \mathrm{d}\mu$;

(2) $\int |X| \mathrm{d}\mu = 0 \Leftrightarrow \mu(X \neq 0) = 0$;

(3) $\left|\int X \mathrm{d}\mu\right| \leqslant \int |X| \mathrm{d}\mu$,

$\left|\int X \mathrm{d}\mu\right| < \infty \Leftrightarrow \int |X| \mathrm{d}\mu < \infty$;

(4) $\int 1 \mathrm{d}\mu = \mu(\Omega)$, $\int cX \mathrm{d}\mu = c\int X \mathrm{d}\mu$, c 为常数；

(5) $\int (X + Y) \mathrm{d}\mu = \int X \mathrm{d}\mu + \int Y \mathrm{d}\mu$，当此三个积分中的任意两个存在.

定义 7.3.4 设 X 为概率空间$\langle \Omega, \mathscr{F}, P \rangle$上的随机变量，且

积分 $\left|\int X\mathrm{d}P\right|<\infty$，则 X 的**期望**(expectation)为 $E(X)=\int X\mathrm{d}P$.

定理 7.3.5 积分变换定理 设 X 为由测度空间 $\langle\Omega,\mathscr{F},\mu\rangle$ 到可测空间 $\langle\Omega',\mathscr{F}'\rangle$ 的可测变换，μ' 为 X 在 $\mathscr{F}'$ 上的导出测度(定理 7.2.13)，g 为 $\langle\Omega',\mathscr{F}'\rangle$ 上的可测函数，则

$$\int g\mathrm{d}\mu' = \int g(X)\mathrm{d}\mu.$$

当且仅当两积分之一存在.

定义 7.3.6 设 F 为 n 元分布函数，P 为它在 n 维博雷尔空间 $\langle\mathbf{R}^n,\mathscr{B}^n\rangle$ 上所决定的概率分布. 对任意 n 维博雷尔函数 f，

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)=\int f\mathrm{d}P,$$

称为**斯蒂尔切斯**(Stieltjes)**积分**.

定义 7.3.7 设 X 为乘积测度空间 $\langle\Omega_1\times\Omega_2,\mathscr{F}_1\times\mathscr{F}_2,\mu_1\times\mu_2\rangle$ 上的可测函数，对 $\forall\ \omega_1\in\Omega_1$，定义 Ω_2 上的函数 $X_{\omega_1}(\omega_2)=X(\omega_1,\omega_2)$，$\forall\ \omega_2\in\Omega_2$，称为 X 在 ω_1 的**切口**(section).

推论 7.3.8 设 X 如(7.3.7)中所定义，X 在 ω_1 的切口 $X_{\omega_1}(\omega_2)$ 是 $\langle\Omega_2,\mathscr{F}_2\rangle$ 上的可测函数.

定理 7.3.9 富比尼(Fubini)定理 设 $\langle\Omega_i,\mathscr{F}_i,\mu_i\rangle$，$i=1,2$ 为两个 σ 有限测度空间，X 为乘积测度空间 $\langle\Omega_1\times\Omega_2,\mathscr{F}_1\times\mathscr{F}_2,\mu_1\times\mu_2\rangle$ 上非负可测(或可积)函数，对 $\forall\ \omega_1\in\Omega_1$，$X_{\omega_1}$ 为 X 在 ω_1 的切口，则积分 $\int X_{\omega_1}\mathrm{d}\mu_2$ 存在，且 $X_1(\omega_1)=\int X_{\omega_1}\mathrm{d}\mu_2$ 作为 $\langle\Omega_1,\mathscr{F}_1,\mu_1\rangle$ 上的函数是非负可测(或 μ_1 可积)的，且

$$\int X\mathrm{d}(\mu_1\times\mu_2)=\int X_1\mathrm{d}\mu_1.$$

7.4 测度的分解

定义 7.4.1 设 X 为测度空间 $\langle\Omega,\mathscr{F},\mu\rangle$ 上积分存在的可测

函数，$\mathscr{F}$ 上的(集合)函数 ν：

$$\nu(A)=\int_A X\mathrm{d}\mu=\int XI_A\mathrm{d}\mu,\quad \forall\ A\in\mathscr{F}.$$

称为 X 关于 μ 的**不定积分**(indefinite integral).

推论 7.4.2 设 X 满足定义(7.4.1)的条件，ν 为 X 关于 μ 的不定积分，则

(1) 对 $A_n\in\mathscr{F}, n\geqslant 1, A_nA_m=\varnothing$，当 $n\neq m$，有

$$\nu\Big(\bigcup_{n=1}^{\infty}A_n\Big)=\sum_{n=1}^{\infty}\nu(A_n);$$

(2) 若 $X\geqslant 0$，则 ν 为 $\langle\Omega,\mathscr{F}\rangle$ 上的测度；

(3) $\mu(A)=0\Rightarrow\nu(A)=0$.

定义 7.4.3 设 μ_1,μ_2 为可测空间 $\langle\Omega,\mathscr{F}\rangle$ 上的两个测度，

(1) 若对 $\forall\ A\in\mathscr{F}$，$\mu_2(A)=0\Rightarrow\mu_1(A)=0$，则称 μ_1 为 μ_2 **连续的**(μ_2-continuous)；

(2) 若存在 $N\in\mathscr{F}$，使得 $\mu_2(N)=0$，$\mu_1(\overline{N})=0$，则称 μ_1 为 μ_2 **奇异的**(μ_2-singular).

由此定义可知，若 μ_1 为 μ_2 奇异的，则 μ_2 为 μ_1 奇异的，即任一测度的全部“质量”集中在另一测度的零测集上.

定理 7.4.4 勒贝格(Lebesgue)**定理** 设 ν 和 μ 为可测空间 $\langle\Omega,\mathscr{F}\rangle$ 上的两个 σ 有限测度，则 ν 可以唯一地分解为两个测度 λ_1,λ_2，即

$$\nu(A)=\lambda_1(A)+\lambda_2(A),\forall\ A\in\mathscr{F}.$$

其中 λ_1 为 μ-奇异的，λ_2 为 μ 连续的.

定理 7.4.5 拉东-尼可丁(Radon-Nikodim)定理 设 ν 和 μ 为可测空间 $\langle\Omega,\mathscr{F}\rangle$ 上的两个 σ 有限测度，则 ν 为 μ 连续的充分必要条件为存在一个 $\langle\Omega,\mathscr{F}\rangle$ 上的非负可测函数 g，使得 ν 为 g 关于 μ 的不定积分，g 在相差一个 μ 零测集的意义下是唯一确定的.

定义 7.4.6 定理 7.4.5 中的 g 称为 ν 关于 μ 的拉东-尼可丁

导数,记为$\frac{d\nu}{d\mu}$,因此

$$\nu(A)=\int_A \frac{d\nu}{d\mu}d\mu, \quad \forall A\in\mathscr{F}.$$

定理 7.4.7 设ν和μ为可测空间$\langle\Omega,\mathscr{F}\rangle$上的两个$\sigma$有限测度,且$\nu$为$\mu$连续,$X$为$\mathscr{F}$可测函数,$\int X d\nu$存在,则

$$\int_A X d\nu=\int_A X\frac{d\nu}{d\mu}d\mu, \forall A\in\mathscr{F}.$$

定理 7.4.8 设ν,μ和λ为可测空间$\langle\Omega,\mathscr{F}\rangle$上的$\sigma$有限测度,且$\nu$为$\mu$连续而$\lambda$为$\nu$连续,则$\lambda$为$\mu$连续,且

$$\frac{d\lambda}{d\mu}=\frac{d\lambda}{d\nu}\frac{d\nu}{d\mu}.$$

定理 7.4.9 设F为分布函数,则

$$F=\lambda_1F_1+\lambda_2F_2+\lambda_3F_3,$$

其中λ_i为非负常数,$i=1,2,3$, $\sum_{i=1}^{3}\lambda_i=1$,$F_1$为纯跳跃分布函数,$F_2$为连续而处处不可导的分布函数,$F_3$为绝对连续分布函数,即存在密度函数$f(x)$,使得

$$F_3(x)=\int_{-\infty}^{x}f(u)du,\forall x\in\mathbf{R}^1.$$

7.5 条件期望与条件概率

定理 7.5.1 设X为概率空间$\langle\Omega,\mathscr{F},P\rangle$上的随机变量,$E(X)$存在,$Z$为$\langle\Omega,\mathscr{F}\rangle$到可测空间$\langle\Omega',\mathscr{F}'\rangle$的可测变换,记$P'$为$Z$在$\mathscr{F}'$上导出的概率测度,则存在$\mathscr{F}'$可测函数$g$,使得

$$\int_{Z^{-1}(B)} X dP=\int_B g dP',\forall B\in\mathscr{F}',$$

且g在相差一个P'零测集的意义下唯一确定.

定义 7.5.2 设 X,Z 和 g 满足定理 7.5.1 中条件，则称 g 为 X 关于 Z 的**条件期望**(conditional expectation)，记为 $E(X|Z)$. 对 $\forall z \in \Omega'$，记 $E(X|Z=z)=g(z)$，称为给定 $Z=z$ 时，X 的**条件期望**.

定义 7.5.3 设 $(\Omega,\mathscr{F},P)$ 为概率空间，$A,B\in\mathscr{F}$，Z 为 $(\Omega,\mathscr{F})$ 到可测空间 $(\Omega',\mathscr{F}')$ 的可测变换，

(1) 定义事件 A 关于 Z 的**条件概率**为

$$P(A|Z) = E(I_A|Z);$$

(2) 定义事件 A 关于事件 B 的条件概率为

$$P(A|B) = E(I_A|I_B).$$

由于条件期望不具有 σ 可加性，一般的条件概率 $P(A|Z)$，对给定的 $Z=z$，作为 $\mathscr{F}$ 上的一个集合函数，不必是一个概率测度.

定理 7.5.4 在定义 7.5.3 中，若 $\Omega=\mathbf{R}^n$，$\mathscr{F}=\mathscr{B}^n$，$n=1,2,\cdots,\infty$，则 $P(A|Z)$ 对给定的 $Z=z$，作为 $\mathscr{F}$ 上的(集合)函数是一个概率测度，称为**正则条件概率**(regular conditional probability).

定理 7.5.5 设 X 和 Y 为概率空间 $(\Omega,\mathscr{F},P)$ 上的两个随机变量，对 $\forall\ y\in\mathbf{R}^1$，存在 $(\mathbf{R}^1,\mathscr{B}^1)$ 上的概率分布 P_y，满足

$$P_y(B) = P(X^{-1}(B)|Y = y),\forall\ B \in \mathscr{B}.$$

$P_y(B)$ 称为**给定 $Y=y$ 时，X 的条件概率分布**.

定理 7.5.6 设 X,Y 为两个随机变量，$F(\cdot|y)$ 为给定 $Y=y$ 时 X 的条件概率分布函数，则

(1) 对任一博雷尔函数 h，有

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x)\mathrm{d}F(x|y);$$

(2) 若 (X,Y) 有联合密度 $f(x,y)$，则

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x)f(x|y)\mathrm{d}x,$$

其中 $f(x|y) = f(x,y)/\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$，称为**给定 $Y=y$ 时，X 的条件概率密度**.

8　随机变量列的几种收敛性

8.1　几种收敛性的定义及强弱关系

以下设 X，$\{X_n, n\geq 1\}$ 为同一概率空间上的随机变量列.

定义 8.1.1　设对任给 $\varepsilon>0$，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0,$$

则称 $\{X_n\}$ **依概率收敛到** X(convergent in probability to X)，记为

$$\lim_{n\to\infty} X_n = X(\text{inP})，\text{或 } X_n \xrightarrow{\text{P}} X.$$

定义 8.1.2　设

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

则称 $\{X_n\}$ **几乎处处收敛到** X (convergent to X almost every where)，记为

$$\lim_{n\to\infty} X_n = X(\text{a. e.})，\text{ 或 } X_n \xrightarrow{\text{a. e.}} X.$$

定义 8.1.3　设对某个实数 $r>0$，有

$$\lim_{n\to\infty} E|X_n - X|^r = 0,$$

则称 $\{X_n\}$ **依 r 阶矩收敛到** X (convergent in moment of order r to X)，记为

$$\lim_{n\to\infty} X_n = X(\text{L}_r)，\text{或 } X_n \xrightarrow{\text{L}_r} X.$$

定义 8.1.4　以 F 和 F_n 分别记 X 和 X_n 的分布函数，以 $C(F)$ 记 F 的连续点的集合，若

$$\lim_{n\to\infty} F_n(x) = F(x), \forall\, x \in C(F).$$

则称 X_n **依分布收敛到** X (convergent in distribution to X)，记为

$$\lim_{n\to\infty} X_n = X(\mathrm{d}), \text{或 } X_n \xrightarrow{\mathrm{d}} X.$$

注 对依分布收敛来说，由于仅与 X 和 X_n 的分布函数有关，因此，X 和 $\{X_n\}$ 不必是同一概率空间上的随机变量.

以上 4 种关于随机变量列的收敛性之间存在着一定的强弱关系.

定理 8.1.5 $X_n \xrightarrow{\mathrm{a.e.}} X \Rightarrow X_n \xrightarrow{\mathrm{P}} X$.

上述定理的逆命题不成立.

例 8.1.6 设概率空间 $(\Omega, \mathscr{F}, P)$：$\Omega=[0,1]$，$\mathscr{F}$ 为 $[0,1]$ 上的博雷尔可测集，P 为 $[0,1]$ 上的勒贝格测度，定义随机变量列：

$$Y_{kj}(\omega) = \begin{cases} k^{\frac{1}{2}}, & \text{当 } \omega \in \left[\dfrac{j-1}{k}, \dfrac{j}{k}\right], \\ 0, & \text{其它}. \end{cases}$$

$j=1,\cdots,k$，$k=1,2,\cdots$，将 $Y_{11}, Y_{21}, Y_{22}, Y_{31}, Y_{32}, Y_{33}, \cdots$ 排成一列，记为 $X_1, X_2, \cdots$，则容易验证

$$X_n \xrightarrow{\mathrm{P}} 0.$$

但 $P(X_n \text{发射})=1$.

定理 8.1.7 $X_n \xrightarrow{L_r} X$，对某个实数 $r>0 \Rightarrow X_n \xrightarrow{\mathrm{P}} X$.

上述定理的逆命题不成立.

例 8.1.8 设 $(\Omega, \mathscr{F}, P)$ 如例 8.1.6 中定义，令

$$X_n(\omega) = \begin{cases} \mathrm{e}^n, & \text{当 } \omega \in \left[1-\dfrac{1}{n}, 1\right], \\ 0, & \text{其它}. \end{cases}$$

则易于验证：$X_n \xrightarrow{\mathrm{P}} 0$，但对任意 $r>0$，有

$$E|X_n - 0|^r = n^{-1}\mathrm{e}^{rn} \to \infty.$$

定理 8.1.9 $X_n \xrightarrow{\mathrm{P}} X \Rightarrow X_n \xrightarrow{\mathrm{d}} X$.

但此定理的逆命题不成立.

例 8.1.10 设 $(\Omega, \mathscr{F}, P)$ 如例 8.1.6 中所定义，令

$$X(\omega) = \begin{cases} 1, & 当\ \omega \in \left[0, \frac{1}{2}\right), \\ 0, & 其它. \end{cases}$$

当 n 为奇数时，令 $X_n = X$；n 为偶数时，$X_n = 1 - X$，则 X_n 与 X 同分布，因而 $X_n \xrightarrow{d} X$，但 $P(|X_{2n} - X| = 1) = 1$.

(8.1.5)～(8.1.10)说明，几乎处处收敛与矩收敛强于依概率收敛，依概率收敛强于依分布收敛，而几乎处处收敛与矩收敛不能相互推出：容易验证例 8.1.6 中 $X_n \xrightarrow{L_r} 0$，但 $P(X_n\ 发散) = 1$. 在例 8.1.8 中 $X_n \xrightarrow{a.e.} 0$，但对$\forall\ r > 0$，$E|X_n - 0|^r \to \infty$.

8.2 积分收敛定理与依矩收敛

定理 8.2.1 单调收敛定理(monotone convergence theorem) 设随机变量列$\{X_n, n \geqslant 1\}$满足

(1) 存在常数 c，使得 $P(X_n \geqslant c, n \geqslant 1) = 1$，

(2) $P(X_n\ 单调非降) = 1$，

则

$$\lim_{n \to \infty} E(X_n) = E(\lim_{n \to \infty} X_n).$$

引理 8.2.2 法图(Fatou)引理　设随机变量列$\{X_n, n \geqslant 1\}$满足

(1) 存在随机变量 Y，使得 $P(X_n \geqslant Y) = 1, n = 1, 2, \cdots$，

(2) $E|Y| < \infty$，

则 $E(\liminf_n X_n)$存在，且

$$E(\liminf_n X_n) \leqslant \liminf_n E(X_n).$$

定理 8.2.3 控制收敛定理(dominated convergence theorem) 设随机变量列$\{X_n, n \geqslant 1\}$满足

(1) $X_n \xrightarrow{P} X$，

(2) 存在随机变量 $Y,EY<\infty$,使得

$$P(|X_n|\leqslant Y)=1,n=1,2,\cdots,$$

则

$$\lim_{n\to\infty}E|X_n-X|=0.$$

以下讨论由 $X_n\xrightarrow{P}X$ 导出 $X_n\xrightarrow{L_r}X$ 的条件.

定义 8.2.4 设$\{X_n,n\geqslant1\}$是随机变量列,若对任给的 $\varepsilon>0$,存在 $\delta>0$,使得对任意事件 $A,P(A)<\delta$,有

$$\sup_n\int_A|X_n|\mathrm{d}P<\varepsilon,$$

且

$$\sup_n E|X_n|<\infty.$$

则称它是**一致可积的**(uniformly integrable).

定理 8.2.5 设$\{X_n,n\geqslant1\}$为随机变量列,则存在 $r>0$,使得 $X_n\xrightarrow{L_r}X$,当且仅当

$$X_n\xrightarrow{P}X.$$

且$\{|X_n|^r,n\geqslant1\}$为一致可积的.

验证一致可积性的一个准则为

定理 8.2.6 随机变量列$\{X_n,n\geqslant1\}$为一致可积的,当且仅当

$$\lim_{a\to\infty}(\sup_n\int_{[|X_n|>a]}|X_n|\mathrm{d}P)=0.$$

利用一致可积的概念可将定理 8.2.1~8.2.3 推广如下:

定理 8.2.7 推广的单调收敛定理 设随机变量列$\{X_n,n\geqslant1\}$满足条件

(1) $P(X_n$ 单调非降$)=1$,

(2) $\{X_n,n\geqslant1\}$为一致可积的,

则

$$\lim_{n\to\infty}E(X_n) = E(\lim_{n\to\infty}X_n) > -\infty.$$

定理 8.2.8 推广的法图引理 设随机变量列$\{X_n, n\geqslant 1\}$满足条件

(1) $\{X_n^-, n\geqslant 1\}$为一致可积的，

(2) $E(\liminf_n X_n)$存在，

则

$$E(\liminf_n X_n) \leqslant \liminf_n E(X_n).$$

定理 8.2.9 推广的控制收敛定理 设随机变量列$\{X_n, n\geqslant 1\}$满足条件

(1) $X_n \xrightarrow{P} X$，

(2) $\{X_n, n\geqslant 1\}$为一致可积，

则

$$\lim_{n\to\infty}E|X_n - X| = 0.$$

定理 8.2.9 实际上是定理 8.2.5 在 $r=1$ 时的特例.

8.3 依概率收敛与几乎处处收敛

8.3.1 0-1 律

概率空间上的某些事件发生的概率只可能取 0 或 1，有关的结论通称为 **0-1 律**(zero-one law).

引理 8.3.1 博雷尔-康特利(Borel-Cantelli)引理 设$\{A_n, n\geqslant 1\}$为事件列，且 $\sum_{n=1}^{\infty}P(A_n) < \infty$，则

$$P(\limsup_n A_n) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}A_m\right) = 0.$$

引理 8.3.1 的含义是：若 $\sum_{n=1}^{\infty}P(A_n) < \infty$，则 $A_n, n = 1,2,\cdots$

至多只能发生有限多个.

定理 8.3.2 博雷尔 0-1 律 设$\{A_n, n\geqslant 1\}$为独立事件列,则

$$P(\limsup A_n) = 0 \text{ 或 } 1,$$

当

$$\sum_{n=1}^{\infty} P(A_n) < \infty \text{ 或 } = \infty.$$

定义 8.3.3 设$\{X_n, n\geqslant 1\}$为随机变量列,定义$\{X_n\}$的**尾 σ 域**(tail σ-field)为

$$\bigcap_{n=1}^{\infty} \sigma(X_j, j \geqslant n),$$

其中$\sigma(X_j, j\geqslant n)$为由$\{X_j, j\geqslant n\}$生成的$\sigma$域.

定义 8.3.4 尾σ域中的事件称为**尾事件**,关于尾σ域可测的函数称为**尾函数**.

定理 8.3.5 科尔莫戈罗夫 0-1 律 设$\{X_n, n\geqslant 1\}$为独立随机变量列,则$\{X_n\}$的尾事件只取概率 0 或 1.

推论 8.3.6 设 X 为独立随机变量列$\{X_n, n\geqslant 1\}$的尾函数,则 $P(X=c)=1$,c 为常数.

定理 8.3.7 设$\{X_n, n\geqslant 1\}$为独立随机变量列,则

(1) $\limsup_n X_n$ 与$\liminf_n X_n$ 均依概率 1 为常数.

(2) 若 $X_n \xrightarrow{\text{a.e.}} X$,则 X 依概率 1 为常数.

8.3.2 依概率收敛与几乎处处收敛

定理 8.3.8 随机变量列$\{X_n, n\geqslant 1\}$依概率收敛到某个随机变量 X,当且仅当

(1) $\{X_n\}$的任一子列中有一个子列几乎处处收敛到同一随机变量 X;

当且仅当

(2) 对$\forall\ \varepsilon>0$,有

$$\lim_{n\to\infty}\sup_{m\geqslant n}P(|X_m - X_n| > \varepsilon) = 0.$$

定理 8.3.9 随机变量列$\{X_n, n\geqslant 1\}$几乎处处收敛到某个随机变量 X,当且仅当

(1) 对任给 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\Big(\bigcup_{m=n}^{\infty}\{|X_m - X| > \varepsilon\}\Big) = 0;$$

当且仅当

(2) $\lim\limits_{n\to\infty}\sup\limits_{m\geqslant n}(|X_m - X|) = 0 \quad (\text{in } P)$.

推论 8.3.10 设 $X, X_n, n\geqslant 1$ 为随机变量,满足:对$\forall\ \varepsilon>0$,有

$$\sum_{n=1}^{\infty}P(|X_n - X| > \varepsilon) < \infty,$$

则 $X_n \xrightarrow{\text{a.e.}} X$.

推论 8.3.11 设$\{X_n, n\geqslant 1\}$为独立随机变量列,则 $X_n \xrightarrow{\text{a.e.}} 0$,当且仅当对任给 $\varepsilon>0$,有

$$\sum_{n=1}^{\infty}P(|X_n| > \varepsilon) < \infty.$$

定理 8.3.12 设$\{X, X_n, n\geqslant 1\}$为随机变量列,g 为连续函数,则

$$X_n \xrightarrow{\text{P}} X \Rightarrow g(X_n) \xrightarrow{\text{P}} g(X),$$

$$X_n \xrightarrow{\text{a.e.}} X \Rightarrow g(X_n) \xrightarrow{\text{a.e.}} g(X).$$

定理 8.3.13 设$\{X_n, n\geqslant 1\}$为独立随机变量列,记 $S_n = \sum\limits_{i=1}^{n}X_i, n\geqslant 1$,则 S_n 几乎处处收敛到某个随机变量 S,当且仅当 S_n 依概率收敛到 S.

8.3.3 级数的几乎处处收敛

定理 8.3.14 设$\{X_n, n\geqslant 1\}$为独立随机变量列, $E(X_n) = 0$,

$E(X_n^2) < \infty, n \geqslant 1$. 若 $\sum_{n=1}^{\infty} E(X_n^2) < \infty$，则

$$\sum_{n=1}^{\infty} X_n \quad \text{a.e. 收敛},$$

且

$$\sum_{n=1}^{\infty} X_n \quad L_2 \quad \text{收敛},$$

此时有

$$E\left(\sum_{n=1}^{\infty} X_n\right)^2 = \sum_{n=1}^{\infty} E(X_n^2).$$

引理 8.3.15 设 $\{X_n, Y_n, n \geqslant 1\}$ 为两个随机变量列，且 $P(\limsup_n \{Y_n \neq X_n\}) = 0$，则

$$\sum_{n=1}^{\infty} X_n \quad \text{a.e. 收敛} \Leftrightarrow \sum_{n=1}^{\infty} Y_n \quad \text{a.e. 收敛}.$$

利用定理 8.3.14 及引理 8.3.15 和引理 8.3.1 容易得到

定理 8.3.16 科尔莫戈罗夫三级数定理 设 $\{X_n, n \geqslant 1\}$ 为独立随机变量列，则 $\sum_{n=1}^{\infty} X_n$ a.e.收敛，当且仅当下列三条件都成立：

(1) $\sum_{n=1}^{\infty} P(|X_n| > 1) < \infty$；

(2) $\sum_{n=1}^{\infty} E(X_n I_{(|X_n| \leqslant 1)})$ 收敛；

(3) $\sum_{n=1}^{\infty} E[X_n^2 I_{(|X_n| \leqslant 1)} - E^2(X_n I_{(|X_n| \leqslant 1)})] < \infty$.

科尔莫戈罗夫三级数定理是推导强大数定律的有力工具.

推论8.3.17 设 $\{X_n, n \geqslant 1\}$ 为独立随机变量列，$E(X_n) = 0$，$n \geqslant 1$，且 $\sum_{n=1}^{\infty} E[|X_n| I_{(|X_n| > 1)} + X_n^2 I_{(|X_n| \leqslant 1)}] < \infty$，则 $\sum_{n=1}^{\infty} X_n$ a.e. 收敛.

推论8.3.18 设 $\{X_n, n \geqslant 1\}$ 为独立随机变量列，$\{a_n, n \geqslant 1\}$ 为

常数列. 若$0<\alpha_n\leqslant 2$,且当$1\leqslant\alpha_n\leqslant 2$时,$E(X_n)=0$, $\sum_{n=1}^{\infty}E|X_n|^{\alpha_n}<\infty$,则$\sum_{n=1}^{\infty}X_n$ a.e. 收敛.

8.4 依分布收敛

随机变量的依分布收敛性只与相应的分布函数列的收敛性有关. 在本节中,分布函数是指一元分布函数.

8.4.1 分布函数列的弱收敛与全收敛

定义8.4.1 设$\{F_n,n\geqslant 1\}$为分布函数列.

(1) 若对F的任一连续点x,有

$$\lim_{n\to\infty}F_n(x)=F(x),$$

则称$\{F_n\}$**弱收敛**(weakly converge)到函数F.

记作$F_n\xrightarrow{\text{w}}F$.

(2) 若$F_n\xrightarrow{\text{w}}F$,且$F(-\infty)=0$,$F(+\infty)=1$,则称$\{F_n\}$**全收敛**(completely converge)到函数F.

记作$F_n\xrightarrow{\text{c}}F$.

推论8.4.2 若$F_n\xrightarrow{\text{w}}F$,则$F$为非负单调非降函数;若$F_n\xrightarrow{\text{c}}F$,且$F$为右连续的,则$F$为分布函数.

推论8.4.3 设$X,X_n,n\geqslant 1$为随机变量,$F,F_n,n\geqslant 1$为相应的分布函数,则

$$X_n\xrightarrow{\text{d}}X\Leftrightarrow F_n\xrightarrow{\text{c}}F.$$

例8.4.4 设$P(X_n=n)=1$,则X_n有分布函数

$$F_n(x)=\begin{cases}0,\text{当 }x<n,\\1,\text{当 }x\geqslant n,\end{cases}\quad n\geqslant 1.$$

显然 $F_n \xrightarrow{w} 0$，此时不存在随机变量 X，使得 $X_n \xrightarrow{d} X$.

此例说明，分布函数列的弱收敛是比随机变量的依分布收敛更广泛的概念.

定理8.4.5 设$\{F_n, n\geqslant 1\}$为分布函数列，则 $F_n \xrightarrow{w} F$，当且仅当存在$\mathbf{R}^1$上的一个稠密子集 D，使得

$$\lim_{n\to\infty} F_n(x) = F(x), \forall\ x \in D.$$

推论8.4.6 设$\{F_n, n\geqslant 1\}$为分布函数列，则 $F_n \xrightarrow{w} F$，当且仅当对 F 的任一有理数连续点 x 有 $\lim\limits_{n\to\infty} F_n(x) = F(x)$.

定理8.4.7 设$\{X_n, n\geqslant 1\}$为随机变量列，c 为常数，则

$$X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c.$$

定理8.4.8 设$\{X_n, Y_n, Z_n \mid n\geqslant 1\}$为三个随机变量列，若 $X_n \xrightarrow{d} X$，$Y_n \xrightarrow{P} a$，$Z_n \xrightarrow{P} b$，其中 X 为随机变量，a，b 为常数，则

$$Y_n X_n + Z_n \xrightarrow{d} ax + b.$$

定理8.4.9 设 $F, G, F_n, n\geqslant 1$为分布函数. 若 $F_n \xrightarrow{c} F$，且存在实数列$\{a_n>0, b_n, n\geqslant 1\}$使得 $F_n(a_n x+b_n) \xrightarrow{c} G(x)$，则存在常数 $a>0$和 b，使得 $G(x)=F(ax+b)$，且$\lim\limits_{n\to\infty} a_n=a$，$\lim\limits_{n\to\infty} b_n=b$.

推论8.4.10 设$\{F_n, G_n, n\geqslant 1\}$为分布函数列，$F$，$G$ 为非退化分布函数，$\{a_n>0, \alpha_n>0, b_n, \beta_n, n\geqslant 1\}$为常数列. 若

$$F_n(a_n x + b_n) \xrightarrow{c} F(x),$$

且

$$F_n(\alpha_n x + \beta_n) \xrightarrow{c} G(x),$$

则 $G(x)=F(ax+b)$，其中

$$a = \lim_{n\to\infty}\left(\frac{\alpha_n}{a_n}\right) > 0,$$

$$b = \lim_{n\to\infty} \frac{(\beta_n - b_n)}{a_n}.$$

引理8.4.11哈利-布雷(Helly-Bray)引理 设$\{F_n,n\geqslant 1\}$为分布函数列，g 为连续实函数，$F_n \xrightarrow{w} F$，则对 F 的任意两个连续点 a 和 b 有

$$\lim_{n\to\infty}\int_a^b g\mathrm{d}F_n = \int_a^b g\mathrm{d}F.$$

定理8.4.12 设$\{F_n,n\geqslant 1\}$为分布函数列，g 为连续函数，则

(1) $F_n \xrightarrow{c} F$，且 g 有界 $\Rightarrow \lim\limits_{n\to\infty}\int_{-\infty}^{\infty} g\mathrm{d}F_n = \int_{-\infty}^{\infty} g\mathrm{d}F$.

(2) $F_n \xrightarrow{w} F$，且 $\lim\limits_{|x|\to\infty} g(x) = 0 \Rightarrow \lim\limits_{n\to\infty}\int_{-\infty}^{\infty} g\mathrm{d}F_n = \int_{-\infty}^{\infty} g\mathrm{d}F$.

定义8.4.13 设$\{F_n,n\geqslant 1\}$为分布函数列，g 为连续函数，若

$$\lim_{a\to\infty}\left(\sup_n \int_{[|x|>a]} |g(x)|\mathrm{d}F_n(x)\right) = 0.$$

则称 g 为关于$\{F_n\}$**一致可积的**.

定理8.4.14 设$\{F_n,n\geqslant 1\}$为分布函数列，g 为连续函数，若 $F_n \xrightarrow{w} F$ 且 g 关于$\{F_n\}$一致可积，则 $\int_{-\infty}^{\infty} |g|\mathrm{d}F < \infty$，且

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} g\mathrm{d}F_n = \int_{-\infty}^{\infty} g\mathrm{d}F.$$

定理8.4.15 设$\{F_n,n\geqslant 1\}$为分布函数列，则

$$F_n \xrightarrow{c} F.$$

当且仅当，对任一有界连续函数 g 有

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} g\mathrm{d}F_n = \int_{-\infty}^{\infty} g\mathrm{d}F.$$

定理8.4.16 设$\{F_n,n\geqslant 1\}$为分布函数列，满足

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} x^k\mathrm{d}F_n(x) = \alpha_k,k = 1,2,\cdots.$$

且$\{\alpha_k,k\geqslant 1\}$唯一地决定一个分布函数 F，使得 $\alpha_k = \int_{-\infty}^{\infty} x^k\mathrm{d}F(x)$，$k = 1,2,\cdots$，则

$$F_n \xrightarrow{c} F.$$

定理8.4.17 设分布函数列$\{F_n, n\geqslant 1\}$全收敛到F，且对F的任一不连续点x有$F_n(x\pm) \longrightarrow F(x\pm)$，则$\{F_n\}$在$\mathbf{R}^1$上一致收敛到$F$.

推论8.4.18 设$F_n \xrightarrow{c} F$，且F连续，则F_n一致收敛到F.

8.4.2 分布函数的拓扑性质

定理8.4.19 任意分布函数的无穷族一定存在一个弱收敛子列.

此定理的结论称为分布函数族的**弱紧性**(weak compactness).

定义8.4.20 设F,G为两个分布函数，它们之间的**距离**为

$$\mathrm{d}[F,G] = \sup_{-\infty<x<+\infty} |F(x) - G(x)|.$$

定理8.4.21 分布函数列$\{F_n\}$全收敛到分布函数F，当且仅当$\lim\limits_{n\to\infty}\mathrm{d}[F_n, F]=0$.

定义8.4.22 若$\{F_n\}$的任一子列有一个全收敛子列，则分布函数的无穷族$\{F_n\}$称为是**全紧的**(completely compact).

定义8.4.23 若

$$\lim_{a\to\infty} \sup_n \{1 - F_n(a) + F_n(-a)\} = 0,$$

则分布函族$\{F_n\}$称为是**紧密的**(tight).

定理8.4.24 分布函数列$\{F_n, n\geqslant 1\}$为全紧的，当且仅当$\{F_n\}$为紧密的.

推论8.4.25 分布函数列$\{F_n, n\geqslant 1\}$全收敛，当且仅当它的任一子列有一个子列全收敛到同一个极限函数.

8.4.3 分布函数列的全收敛与特征函数列的收敛性

定理8.4.26 设$\{F_n, n\geqslant 1\}$为分布函数列，相应的特征函数列

为$\{\varphi_n, n\geqslant 1\}$.

(1) $F_n \xrightarrow{c} F \Rightarrow$ 对$\forall\ T>0$,

$$\lim_{n\to\infty}\varphi_n(t) = \varphi(t) \quad 在\ |t|\leqslant T\ 上一致,$$

其中φ为F的特征函数.

(2) 若$\lim\limits_{n\to\infty}\varphi_n=\varphi$, φ在O点连续,则φ为某个分布函数F的特征函数,且

$$F_n \xrightarrow{c} F.$$

推论8.4.27 设$F, F_n, n\geqslant 1$为分布函数,$\varphi, \varphi_n, n\geqslant 1$为相应的特征函数,则

$$F_n \xrightarrow{c} F \Leftrightarrow \varphi_n \longrightarrow \varphi.$$

定理8.4.28 设$\{X_n, n\geqslant 1\}$为独立随机变量列,$S_n = \sum\limits_{i=1}^{n} X_i$, $n = 1,2,\cdots$,则S_n依分布收敛,当且仅当存在随机变量S,使得$S_n \xrightarrow{\text{a.e.}} S$.

9 大数定律

9.1 独立同分布随机变量列的大数定律

在 1.2.2 中曾介绍了频率稳定性这一事实，即当多次互不影响地重复某一随机试验时，任一指定事件 A 发生的频率稳定地趋近于某一定值. 这个定值可解释为事件 A 发生的概率. 用随机变量的形式表示，记

$$X_i = \begin{cases} 1, & \text{当第 } i \text{ 次试验 } A \text{ 发生}, \\ 0, & \text{当第 } i \text{ 次试验 } A \text{ 不发生}, \end{cases} \tag{9.1}$$

$$i = 1, \cdots, n, \cdots.$$

$$S_n = \sum_{i=1}^{n} X_i, \quad n = 1, 2, \cdots, \tag{9.2}$$

则 S_n/n 为前 n 次试验中 A 发生的频率. 因此 S_n/n 稳定地趋向于某一定值 p. 而这个 p 就应该是 A 发生的概率.

伯努利在 1713 年首先给出了频率稳定性的一种数学表示方法，并证明了结论. 这是一系列**大数定律**(law of large number)的结论中的第一个.

定理 9.1.1 伯努利大数定律　在一个"成功"概率为 p 的伯努利试验系列中，$X_n, S_n, n \geqslant 1$ 如式(9.1)和(9.2)所定义，则对任给的正数 ε 和 δ，存在正整数 N，当 $n > N$ 时，有

$$P\left(\left| \frac{S_n}{n} - p \right| \geqslant \varepsilon \right) < \delta,$$

或等价地

$$P\left(\left| \frac{S_n}{n} - p \right| < \varepsilon \right) \geqslant 1 - \delta.$$

伯努利大数定律实际上是说$\frac{S_n}{n}\xrightarrow{\text{P}}p$.这就既给出了频率稳定性的一种数学表达形式又肯定了它的真实性.

博雷尔研究了频率稳定性的另一种数学表达形式,并得出了比伯努利大数定律更强的结论.

定理 9.1.2 博雷尔强大数定律 在一个"成功"概率为p的无穷伯努利试验序列中,$X_n,S_n,n\geqslant 1$如(9.1)和(9.2)所定义,则

$$P\left\{\lim_{n\to\infty}\frac{S_n}{n}=p\right\}=1.$$

博雷尔大数定律给出了$\frac{S_n}{n}\xrightarrow{\text{a.e.}}p$的结论.由于a.e.收敛是比依概率收敛更强(见8.1.5).因此这一结果被称为博雷尔**强大数定律**(strong law of large number).

以上两个定理是关于独立0-1随机变量列的部分和S_n与n之比的收敛性的结论.推广之,就得到一般的大数定律.

定义 9.1.3 设$\{X_n,n\geqslant 1\}$为随机变量列,部分和$S_n=\sum_{i=1}^{n}X_i,n=1,2,\cdots$,若存在常数列$\{a_n,n\geqslant 1\}$,使得

(1) $\frac{S_n}{n}-a_n\xrightarrow{\text{P}}0$,则称**大数定律成立**.

(2) $\frac{S_n}{n}-a_n\xrightarrow{\text{a.e.}}0$,则称**强大数定律成立**.

显然,强大数定律成立则大数定律成立.因此,大数定律又称为**弱大数定律**(weak law of large number).

下面介绍对独立同分布随机变量列大数定律成立的一些结论.

定理 9.1.4 科尔莫戈罗夫强大数定律 设$\{X_n,n\geqslant 1\}$为独立同分布随机变量列,$S_n=\sum_{i=1}^{n}X_i$, $n=1,2,\cdots$,则存在常数列$\{a_n\}$使得$\frac{S_n}{n}-a_n\xrightarrow{\text{a.e.}}0$,当且仅当$|EX_1|<\infty$,此时可取$a_n\equiv EX_1,n=$

1,2,….

科尔莫戈罗夫强大数定律的一个直接应用是证明格里文科(Гливенко)定理.

定理 9.1.5 设$\{X_n,n\geqslant 1\}$为独立同分布随机变量列，$S_n=\sum_{i=1}^{n}X_i, n=1,2,\cdots$. 若$EX_1=\pm\infty$，则

$$\frac{S_n}{n}\xrightarrow{\text{a.e.}}\pm\infty.$$

对弱大数定律，一阶矩的存在是过强的条件.

引理 9.1.6 随机变量X若有有限期望($|EX|<\infty$)，则

$$\lim_{n\to\infty}nP(|X|>n)=0.$$

定理 9.1.7 费勒(Feller)定理 设$\{X_n,n\geqslant 1\}$为独立同分布随机变量列，$S_n=\sum_{i=1}^{n}X_i$，$n=1,2,\cdots$，则存在常数列$\{a_n,n\geqslant 1\}$使得$(S_n/n)-a_n\xrightarrow{\text{P}}0$，当且仅当$\lim_{n\to\infty}nP(|X_1|>n)=0$，此时可取$a_n=E(X_1I_{(|X_1|\leqslant n)}),n=1,2,\cdots$.

当部分和S_n取随机下标时，有

定理 9.1.8 设$\{X_n,n\geqslant 1\}$为独立同分布随机变量列，$nP(|X_1|>n)\to 0$，$\{T_n,n\geqslant 1\}$为一正整值随机变量列，$T_n/n\xrightarrow{\text{P}}c$，$c$为正常数，定义$S_{T_n}=\sum_{i=1}^{T_n}X_i$，则存在常数列$\{a_n\}$，使得

$$(S_{T_n}/T_n)-a_n\xrightarrow{\text{P}}0,$$

其中a_n可取$E(X_1I_{(|X_1|\leqslant n)}),n=1,2,\cdots$.

定理 9.1.9 设$\{X_n,n\geqslant 1\}$为独立同分布随机变量列，$|EX_1|<\infty$，$\{T_n,n\geqslant 1\}$为一正整值随机变量列，$T_n\xrightarrow{\text{a.e.}}\infty$，则

$$S_{T_n}/T_n\xrightarrow{\text{a.e.}}EX_1.$$

关于独立不同分布随机变量列的大数定律在节 9.2 中叙述，它是稳定性的一个特例.

9.2 随机变量和的稳定性

定义 9.2.1 设$\{T_n,n\geqslant 1\}$为随机变量列，若存在常数列$\{b_n,a_n,n\geqslant 1\}$，满足$(T_n-b_n)/a_n \xrightarrow{\text{a.e.}} 0$（或$(T_n-b_n)/a_n \xrightarrow{P} 0$），则称$\{T_n\}$对**中心化常数**$\{b_n\}$与标准化常数$\{a_n\}$具有**强稳定性**（或**弱稳定性**）.

稳定性(stability)是一个比大数定律更一般的概念. 从定义中可见，稳定性问题与中心化常数和标准化常数的选取有关.

定理 9.2.2 设$\{T_n,n\geqslant 1\}$为随机变量列，$m(T_n)$为T_n的中位数，$n=1,2,\cdots$，则$\{T_n\}$关于某中心化常数$\{b_n\}$和标准化常数$\{a_n\}$具有强稳定性（或弱稳定性），当且仅当$\{T_n\}$关于中心化常数$\{m(T_n)\}$和标准化常数$\{a_n\}$具有强稳定性（或弱稳定性），且此时必有

$$\lim_{n\to\infty}(b_n - m(T_n))/a_n = 0.$$

此定理说明，在中心化常数的选取中，T_n的中位数$m(T_n)$是当然的候选者. 另一个重要事实是：在稳定性问题中，总可假定随机变量T_n为对称的.

定义 9.2.3 设T为随机变量，T'是与T相互独立且同分布的随机变量，随机变量$T^s = T - T'$，称为T的**对称化**(symmetrization).

定理 9.2.4 设$\{T_n,n\geqslant 1\}$为随机变量列，$\{T_n^s,n\geqslant 1\}$为$\{T_n\}$的对称化. 则$\{T_n\}$关于中心化常数列$\{b_n\}$和标准化常数列$\{a_n\}$有强（或弱）稳定性，当且仅当$\{T_n^s,n\geqslant 1\}$关于中心化常数 0 和标准化常数列$\{a_n\}$有强（或弱）稳定性.

关于标准化常数，首先应指出它的存在性是有保证的. 设 $\{T_n, n\geqslant 1\}$ 为任一随机变量列，由定理 9.2.4，不妨设 $\{T_n, n\geqslant 1\}$ 为对称随机变量列，取 $a_n>0, a_n\to\infty$，使得

$$P\left\{|T_n| \geqslant \frac{a_n}{n}\right\} \leqslant 2^{-n}, \quad n = 1,2,\cdots.$$

由引理 8.3.1 易知，$T_n/a_n \xrightarrow{\text{a.e.}} 0$.

因此，关键的问题是要选取适当的中心化常数列 $\{a_n\}$，使得 a_n 趋向于无穷的速度尽可能缓慢. 以下集中讨论独立随机变量和的稳定性结果. 首先，对独立同分布随机变量的和有

定理 9.2.5 设 $\{X_n, n\geqslant 1\}$ 为独立同分布随机变量列，$S_n = \sum_{i=1}^{n} X_i, n = 1,2,\cdots$，则存在 $r\in(0,2)$ 及常数列 $\{b_n\}$，使得 $(S_n - b_n)/n^{\frac{1}{r}} \xrightarrow{\text{a.e.}} 0$ 的充分必要条件是 $E|X_1|^r<\infty$. 此时，当 $0<r<1$，可取 $b_n\equiv 0$；当 $1\leqslant r<2$，可取 $b_n = nEX_1$.

在此定理中，当 $r=1$ 时，即为科尔莫戈罗夫强大数定律，当 $1<r<2$ 时，给出了比强大数定律更为精细的结果.

定理 9.2.6 设 $\{X_n, n\geqslant 1\}$ 为独立同分布随机变量列，$S_n = \sum_{i=1}^{n} X_i, n = 1,2,\cdots$. 设 $0<r<1$ 或 $1\leqslant r<2$，且 $EX_1=0$，则下列两个结论必成立一个：

(1) $P\left(\lim\limits_{n\to\infty}\left(S_n/n^{\frac{1}{r}}\right) = \lim\limits_{n\to\infty}\left(X_n/n^{\frac{1}{r}}\right) = 0\right) = 1$.

(2) $P\left(\limsup\limits_{n}\left(|S_n|/n^{\frac{1}{r}}\right) = \limsup\limits_{n}\left(|X_n|/n^{\frac{1}{r}}\right) = \infty\right) = 1$.

此定理说明，对独立同分布随机变量列 $\{X_n\}$，其部分和 S_n 与 X_n 有相同的振荡幅度.

定理 9.2.7 设 $\{X_n, n\geqslant 1\}$ 为独立同分布随机变量列，$0<E(X_1^2)<\infty, S_n = \sum_{i=1}^{n} X_i, n=1,2,\cdots$，则对任一实数列

$\{b_n, n\geqslant 1\}$有

$$P\left(\limsup_n |S_n - b_n|/n^{\frac{1}{2}} = \infty\right) = 1.$$

此定理指出，独立同分布随机变量的部分和 S_n 围绕任一中心化常数列振荡的幅度的阶总要高于 $n^{\frac{1}{2}}$，更精细的结果是重对数律（定理 9.3.1）.

关于独立不同分布随机变量列的强稳定性有

定理 9.2.8 设$\{X_n, n\geqslant 1\}$为独立随机变量列，$S_n = \sum_{i=1}^{n} X_i$，$n = 1,2,\cdots$；$\{a_n, \alpha_n, n\geqslant 1\}$ 为两个正常数列，$a_n \uparrow \infty$，$0<\alpha_n\leqslant 2$，且当 $1\leqslant\alpha_n<2$ 时，$EX_n=0$. 则当

$$\sum_{n=1}^{\infty} E|X_n|^{\alpha_n}/a_n^{\alpha_n} < \infty$$

时，有

$$S_n/a_n \xrightarrow{\text{a.e.}} 0.$$

推论 9.2.9 科尔莫戈罗夫推论　设$\{X_n, n\geqslant 1\}$为独立随机变量列，$S_n = \sum_{i=1}^{n} X_i$，则当

$$\sum_{n=1}^{\infty} \sigma^2(X_n)/n^2 < \infty$$

时，有

$$(S_n - ES_n)/n \xrightarrow{\text{a.e.}} 0.$$

关于独立随机变量的弱稳定性有

定理 9.2.10 设$\{X_n, n\geqslant 1\}$为独立随机变量列，$S_n = \sum_{i=1}^{n} X_i$，$n = 1,2,\cdots$；$\{a_n, n\geqslant 1\}$ 为正常数列，$a_n \uparrow \infty$，则 $S_n/a_n \xrightarrow{\text{P}} 0$，当且仅当下列三条件同时成立.

(1) $\lim\limits_{n\to\infty} \sum_{i=1}^{n} P(|X_i| \geqslant a_n) = 0.$

(2) $\lim_{n\to\infty}\frac{1}{a_n}\sum_{i=1}^{n}E(X_iI_{[|X_i|<a_n]})=0$.

(3) $\lim_{n\to\infty}\frac{1}{a_n^2}\sum_{i=1}^{n}\sigma^2(X_iI_{[|X_i|<a_n]})=0$.

9.3 重对数律

设$\{X_n,n\geqslant1\}$为独立同分布随机变量列,$0<E(X_1^2)<\infty$,则根据定理 **9.2.5** 知,部分和 S_n 围绕中心化常数振荡的幅度的阶小于 $n^{\frac{1}{r}}$,$0<r<2$. 而定理 9.2.7 又指出这种振荡幅度的阶总大于 $n^{\frac{1}{2}}$. 进一步的问题是:这种振荡幅度的"精确值"是什么?一些称为**重对数律**(law of iterated logarithm)的结论回答了这个问题.

定理 9.3.1 哈特曼-文特纳尔(Hartman-Wintner)重对数律 设$\{X_n,n\geqslant1\}$为独立同分布随机变量列,

$$E(X_1)=0,0<E(X_1^2)<\infty,S_n=\sum_{i=1}^{n}X_i,n=1,2,\cdots,$$

则

$$P\left(\limsup_n\left[S_n/(n\sigma^2\log\log(n\sigma^2))^{\frac{1}{2}}\right]=\sqrt{2}\right)=1,$$

其中 $\sigma^2=E(X_1^2)$,log log 是双重自然对数.

推论 9.3.2 在上述定理的条件和记号下有

$$P\left(\liminf_n\left[S_n/(n\sigma^2\log\log(n\sigma^2))^{\frac{1}{2}}\right]=-\sqrt{2}\right)=1.$$

定理 9.3.1 重对数律说明,对非退化的独立同分布随机变量列$\{X_n\}$,只要二阶矩 $\sigma^2=EX_1^2$ 有限(不妨设 $\sigma^2=1$),则部分和 S_n 围绕中心化常数振荡幅度为$(2n\log\log n)^{\frac{1}{2}}$. 更确切地说,对任意小的正数 δ,S_n-ES_n 与$(2n\log\log n)^{\frac{1}{2}}$之比最多只有有限多次大于 $1+\delta$或小于$-(1+\delta)$,而有无穷多次大于 $1-\delta$,有无穷多次小于$-(1-\delta)$.

对独立不同分布的随机变量列有

定理 9.3.3 科尔莫戈罗夫重对数律 设$\{X_n, n \geqslant 1\}$为独立随机变量列，$E(X_n)=0, \sigma_n^2=E(X_n^2)<\infty, n=1,2,\cdots$. 记 $S_n^2=\sum_{i=1}^{n} \sigma_i^2, n=1,2,\cdots$. 设$S_n^2 \to \infty$，且存在正常数列$\{K_n\}$，满足$K_n \to 0$，使得

$$P(|X_n| \leqslant K_n S_n (\log \log S_n)^{-\frac{1}{2}})=1, n=1,2,\cdots,$$

则

$$P(\limsup_n [S_n/(2S_n^2 \log \log S_n^2)^{\frac{1}{2}}]=1)=1,$$

$$P(\liminf_n [S_n/(2S_n^2 \log \log S_n^2)^{\frac{1}{2}}]=-1)=1.$$

定理 9.3.4 设$\{X_n, n \geqslant 1\}$为独立随机变量列，$E(X_n)=0$，$E(X_n^2)<\infty$，$S_n^2=\sum_{i=1}^{n} E(X_i^2) \to \infty$，且下述三条件成立：

(1) $\sum_{n=1}^{\infty} P(|X_n|>\delta S_n (\log \log S_n^2)^{\frac{1}{2}})<\infty$，对某个 $\delta>0$，

(2) $S_n^{-2} \sum_{i=1}^{n} E\left[X_i^2 I_{[|X_i|>\varepsilon S_i (\log \log S_i^2)^{-\frac{1}{2}}]}\right] \to 0$，对 $\forall\ \varepsilon>0$，

(3) $\sum_{n=1}^{\infty}[S_n^2(\log \log S_n^2)]^{-1} E\left[X_n^2 I_{[\varepsilon S_n (\log \log S_n^2)^{-\frac{1}{2}}<|X_n|<\delta S_n (\log \log S_n^2)^{\frac{1}{2}}]}\right]$
$<\infty$， 对 $\forall\ \varepsilon>0$，

则

$$P\left\{\limsup_n \frac{S_n}{S_n \sqrt{\log \log S_n^2}}=\sqrt{2}\right\}=1,$$

$$P\left\{\liminf_n \frac{S_n}{S_n \sqrt{\log \log S_n^2}}=-\sqrt{2}\right\}=1.$$

10 中心极限定理

10.1 古典中心极限定理

10.1.1 伯努利试验序列的中心极限定理

在概率论发展的早期，伯努利试验序列是重要的研究对象，而二项分布作为 n 重伯努利试验序列中“成功”次数的分布，它的性质得到了仔细的研究. 继伯努利于 1713 年得到了以他的名字命名的大数定律（9.1.1）之后，棣莫弗（De Moivre）和拉普拉斯（Laplace）从近似计算二项分布的概率出发，得到了第一个**中心极限定理**（central limit theorem）.

定理 10.1.1 棣莫弗-拉普拉斯中心极限定理　设随机变量 $S_n \sim B(n,p)$，$0<p<1$，$n=1,2,\cdots$，则对任何实数 x 有

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{np(1-p)}} \leqslant x\right) = \Phi(x),$$

其中 $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du$，为标准正态分布的分布函数.

定理 10.1.1 又称为**积分极限定理**.

定理 10.1.2 局部中心极限定理　设随机变量 $S_n \sim B(n,p)$，$0<p<1$，$n=1,2,\cdots$，设 $k=k(n)$，$x_k=(k-np)/[np(1-p)]^{\frac{1}{2}}$，则

(1) 若当 $n\to\infty$ 时，x_k 恒属于有限区间 $[a,b]$，就有

$$\lim_{n\to\infty}\frac{P(S_n=k)}{[np(1-p)]^{\frac{1}{2}}\varphi(x_k)} = 1 \quad 一致,$$

其中 $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，为标准正态分布的密度.

(2) 若 $x_k n^{-\frac{1}{6}} \to 0$，当 $n\to\infty$，则

$$\left|\frac{P(S_n = k)}{[np(1-p)]^{\frac{1}{2}}\varphi(x_k)} - 1\right| < \frac{A}{n} + \frac{B|x|^3}{\sqrt{n}} + \frac{C|x|}{\sqrt{n}},$$

其中 A，B 和 C 为与 n 无关的常数.

在定理 10.1.1 中，考虑的是标准化的 S_n 在固定端点的区间 $(-\infty, x]$内的概率的极限，当区间端点 x 也随 n 变化且趋于$+\infty$时，有

定理 10.1.3 设随机变量 $S_n \sim B(n,p)$，$0<p<1$，$n=1,2,\cdots$，$\{x_n\}$为常数列，当 $n\to\infty$时，$x_n\to\infty$，$x_n n^{-\frac{1}{6}}\to 0$，则

$$\lim_{n\to\infty}\frac{P\left(\dfrac{S_n - np}{\sqrt{np(1-p)}} \geqslant x_n\right)}{1-\Phi(x_n)} = 1.$$

此定理说明，当 x_n 趋向于$+\infty$的速度低于 $n^{\frac{1}{6}}$的阶时，标准化的 S_n 超过 x_n 的概率与$(1-\Phi(x_n))$是等价的无穷小. 进一步还可将无穷小量$(1-\Phi(x_n))$表为标准正态密度 φ 的形式.

引理 10.1.4 设 $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，$\Phi(x) = \displaystyle\int_{-\infty}^{x}\varphi(u)du$，则

(1) 对 $x>0$，有

$$\frac{x}{1+x^2}\varphi(x) < 1-\Phi(x) < \frac{1}{x}\varphi(x).$$

(2) $\displaystyle\lim_{x\to\infty}\frac{(1-\Phi(x))x}{\varphi(x)} = 1.$

由(10.1.3)和(10.1.4)便有

推论 10.1.5 设随机变量 $S_n\sim B(n,p)$，$0<p<1$，$n=1,2,\cdots$，常数列$\{x_n\}$满足：$x_n\to\infty$，$x_n n^{-\frac{1}{6}}\to 0$，当 $n\to\infty$，则 $P(S_n>np+x_n\sqrt{np(1-p)})$与$\dfrac{1}{x_n}\varphi(x_n)$为等价无穷小.

泊松从另一个角度考虑了二项分布的极限. 他研究一系列二项分布 $B(n,p_n)$，其中 p_n 随 n 变化的情形.

定理 10.1.6 泊松极限定理 设随机变量$S_n \sim B(n, p_n)$, $n=1,2,\cdots$. 若存在正常数 λ,使得 $np_n \to \lambda$,当 $n\to\infty$,则对给定的正整数 k,有

$$\lim_{n\to\infty} P(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

棣莫费-拉普拉斯中心极限定理和泊松极限定理在计算二项概率分布中有重要应用. 另外,它们又是进行一系列更为深刻的研究的开端.

10.1.2 古典中心极限问题

将定理 10.1.1 的形式改变一下,可得到一个等价的表述:设$\{X_n, n\geqslant 1\}$为独立同分布随机变量列,$P(X_1=1)=p=1-P(X_1=0)$,则 $S_n = \sum_{i=1}^{n} X_i$ 满足

$$(S_n - ES_n)/\sigma(S_n) \xrightarrow{d} N(0,1). \tag{10.1}$$

推广到一般的随机变量列,就得到一般的中心极限定理成立的定义.

定义 10.1.7 设$\{S_n, n\geqslant 1\}$为随机变量列,若式(10.1)式成立,则称**对随机变量列$\{S_n\}$中心极限定理成立**.

所谓古典中心极限问题,就是限于研究独立随机变量列的部分和的中心极限定理. 在定义 10.1.7 中,ES_n 和 $\sigma(S_n)$分别称为中心化常数和标准化常数. 由于极限分布限定是 $N(0,1)$,对其它的中心化常数和标准化常数有比较简单的结论.

推论 10.1.8 设$\{S_n, n\geqslant 1\}$为随机变量,$ES_n^2<\infty$,若存在常数列$\{A_n, n\geqslant 1\}$和$\{B_n>0, n\geqslant 1\}$,

$$(S_n - A_n)/B_n \xrightarrow{d} N(0,1),$$

则有

$$\lim_{n\to\infty}\frac{\sigma(S_n)}{B_n}=1,\text{且}\lim_{n\to\infty}\frac{A_n-ES_n}{B_n}=0.$$

当 S_n 是独立同分布随机变量列的部分和时，有

定理 10.1.9 设$\{X_n,n\geqslant1\}$为独立同分布随机变量列，记 $\mu=EX_1,\sigma^2=\sigma^2(X_1)$，有 $0<\sigma^2<\infty$, $S_n=\sum_{i=1}^{n}X_i,n=1,2,\cdots$，则

$$(S_n-n\mu)/\sqrt{n\sigma^2}\xrightarrow{\text{d}}N(0,1).$$

定理 10.1.1 是此定理的一个特例. 对独立不同分布的随机变量列有

定理 10.1.10 李雅普诺夫(Liapunov)定理 设$\{X_n,n\geqslant1\}$为独立随机变量列，存在 $\delta>0$，使得 $E|X_n|^{2+\delta}<\infty,n=1,2,\cdots$. 记 $S_n^2=\sum_{i=1}^{n}\sigma^2(X_i),\Gamma_n^{2+\delta}=\sum_{i=1}^{n}E|X_i-EX_i|^{2+\delta}$. 若 $\Gamma_n/S_n\to0$，当 $n\to\infty$，则

$$(S_n-ES_n)/S_n\xrightarrow{\text{d}}N(0,1).$$

推论 10.1.11 设$\{X_n,n\geqslant1\}$为独立随机变量列，存在常数 $c>0$，使得 $P(|X_n|\leqslant c)=1,n=1,2,\cdots$. 记 $S_n=\sum_{i=1}^{n}X_i,S_n^2=\sum_{i=1}^{n}\sigma^2(X_i)$，则若 $S_n\to\infty$，就有

$$(S_n-ES_n)/S_n\xrightarrow{\text{d}}N(0,1).$$

比李雅普诺夫定理更一般的结果是如下定理.

定理 10.1.12 贝利-埃森(Berry-Esseen)定理 设$\{X_n,n\geqslant1\}$为独立随机变量列，存在 $\delta\in(0,1]$，使得 $E|X_n|^{2+\delta}<\infty$, S_n^2 和 $\Gamma_n^{2+\delta}$(如定理 10.1.10 中所定义)，则存在一个仅依赖于 δ 的正常

数 c_δ，使得

$$\sup_{-\infty<x<\infty}\left|P\left(\frac{S_n-ES_n}{S_n}<x\right)-\Phi(x)\right|\leqslant c_\delta\left(\frac{\Gamma_n}{S_n}\right)^{2+\delta}.$$

此不等式的右端常称为贝利-埃森界，可用来估计渐近正态的一致速度. 特别当 $\{X_n\}$ 为独立同分布且 $\delta=1$ 时，贝利-埃森界是 $n^{-\frac{1}{2}}$ 阶的.

李雅普诺夫定理的一个主要缺点是要求高于二阶的矩存在且有限. 林德贝格(Lindeberg)首先给出了古典中心极限定理成立的一个重要的充分条件.

引理 10.1.13 设 $\{X_n,n\geqslant1\}$ 为独立随机变量列，$EX_n=0$，$E(X_n^2)<\infty$，$S_n^2=\sum_{i=1}^n E(X_i^2)$，$n=1,\cdots$，记 X_n 的分布函数为 F_n，则下列二条件等价：

(1) $\lim\limits_{n\to\infty}S_n^{-2}\sum_{j=1}^n\int_{[|x|>\varepsilon S_j]}x^2\mathrm{d}F_j(x)=0$，　对 $\forall\ \varepsilon>0$，

(2) $\lim\limits_{n\to\infty}S_n^{-2}\sum_{j=1}^n\int_{[|x|>\varepsilon S_n]}x^2\mathrm{d}F_j(x)=0$，　对 $\forall\ \varepsilon>0$.

定义 10.1.14 设独立随机变量列 $\{X_n,n\geqslant1\}$ 满足引理 10.1.13 中的条件，则称**林德贝格(Lindeberg)条件**成立.

定理 10.1.15 设独立随机变量列 $\{X_n,n\geqslant1\}$ 满足林德贝格条件，则中心极限定理成立.

费勒证明了：对独立随机变量列作一些合理的限制后，林德贝格条件还是中心极限定理成立的一个必要条件(定理 10.1.19).

引理 10.1.16 设 $\{X_n,n\geqslant1\}$ 为独立随机变量列，$\sigma^2(X_n)<\infty$，$n=1,2,\cdots$，记 $S_n^2=\sum_{i=1}^n\sigma^2(X_i)$，$n=1,2,\cdots$，下列两条件等价：

(1) $\lim\limits_{n\to\infty}S_n^2=\infty$，且 $\lim\limits_{n\to\infty}\left(\max\limits_{1\leqslant j\leqslant n}\frac{\sigma^2(X_j)}{S_n^2}\right)=0$，

(2) $\lim_{n\to\infty} S_n^2 = \infty$，且$\lim_{n\to\infty}\left(\frac{\sigma^2(X_n)}{S_n^2}\right) = 0$.

定义 10.1.17 设独立随机变量列$\{X_n, n\geqslant 1\}$满足引理10.1.16中的条件，则称**一致渐近可忽略**(uniformly asymptotic negligible)条件成立.

一致渐近可忽略条件的含义是：当n充分大时，部分和S_n的方差应充分大，而每个X_j在部分和S_n中所占的比重充分小.这就避免了一些最平凡的情形，使得所论内容不至流于空泛.例如，容易证明

定理 10.1.18 对独立随机变量列$\{X_n, n\geqslant 1\}$若$S_n^2 = \sum_{i=1}^{n}\sigma^2(X_i) \to c$，当$n\to\infty$，$c$为有限值，且对部分和$S_n$中心极限定理成立，则每个$X_i$都服从正态分布.

又例如设$X_1\sim N(0,1)$，而$X_n\equiv 0, n\geqslant 2$，则中心极限定理自然成立，而这结论过于平凡.

定理 10.1.19 费勒定理 设独立随机变量列$\{X_n, n\geqslant 1\}$满足一致渐近可忽略条件，且对部分和S_n中心极限定理成立，则林德贝格条件成立.

林德贝格条件中自然蕴含了$S_n^2\to\infty$.由于$\int_{[|x|>\epsilon S_n]} x^2 \mathrm{d}F_j(x)$反映了$X_j$分布的"尾巴"，因此，林德贝格条件的一个直观解释无疑是说，只有$\{X_n\}$的分布函数族中有一种普遍的"轻尾"效应，中心极限定理才有可能成立.

例 10.1.20 设$\{X_n, n\geqslant 1\}$为独立随机变量列，X_n有密度

$$f_n(x) = \frac{2+n^{-\frac{1}{2}}}{2}|x|^{-(3+n^{-\frac{1}{2}})},\ -\infty < x < \infty,$$

则容易验证，一致渐近可忽略条件成立，而林德贝格条件不成立，因而中心极限定理不成立.

10.2 普遍中心极限问题与无穷可分分布族

10.2.1 普遍中心极限问题与无穷可分分布族

在古典中心极限问题中研究独立随机变量列的部分和经过适当的中心化和标准化后依分布收敛到标准正态分布的条件. 该问题可以从两个方面推广. 首先，泊松极限定理启发人们考虑一列独立随机变量组 $\{X_{n1},\cdots,X_{nk_n}\}$，$n=1,2,\cdots$（不同的随机变量组之间不必有直接的联系），求组和 $S_n=\sum_{i=1}^{k_n}X_{ni}$ 依分布收敛到某一分布 F 的问题. 其次，极限分布也不必是正态分布，这就是普遍中心极限问题.

为使问题的讨论有意义，规定独立随机变量组列有下述的性质

定义 10.2.1 如果下述条件成立：

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant k_n}P(|X_{nk}|\geqslant\varepsilon)=0,\text{对}\ \forall\ \varepsilon>0,$$

称独立随机变量组列 $\{X_{n1},\cdots,X_{nk_n}\}$，$n=1,2,\cdots$，$\lim\limits_{n\to\infty}k_n=\infty$ 满足**一致渐近可忽略条件**.

如果不规定一致渐近可忽略条件，则任一分布函数 F 均成为可能的极限分布. 例如取 $X_{n1}\sim F$，而 $X_{nk}\equiv 0$，$k=2,\cdots,k_n$，$n=1,2,\cdots$，则 $S_n\equiv X_{n1}\sim F$. 这样，问题的讨论便失去意义.

定义 10.2.2 若对任给正整数 n，存在一个分布函数 F_n，使得 F 是 F_n 的 n 重卷积，则分布函数 F 称为是**无穷可分的**(infinitely divisible).

推论 10.2.3 设 F 为无穷可分分布函数，则

(1) 对任给正整数 n，存在特征函数 φ_n，使得 F 的特征函数 $\varphi=(\varphi_n)^n$，

(2) 对任给正整数 n，可构造 n 个独立同分布的随机变量 $X_{n1},\cdots,X_{nn}$，使得 $\sum_{k=1}^{n} X_{kn} \sim F$.

定理 10.2.4 设独立随机变量组列 $\{X_{n1},\cdots,X_{nk_n}\}$，$n=1,2,\cdots$ 满足条件 $\lim\limits_{n\to\infty} k_n=\infty$ 及一致渐近可忽略条件，则 $S_n=\sum_{k=1}^{k_n} X_{nk}$ 的一切可能的极限分布族为无穷可分分布族.

下面讨论无穷可分分布族的特征.

定理 10.2.5 泊松型分布的有限卷积的极限分布类即为无穷可分分布族.

定理 10.2.6 列维-辛钦(Levy-Хинчин)定理 分布函数 F 为无穷可分的，当且仅当它的特征函数 φ 具有形式

$$\varphi(t)=\exp\left\{\mathrm{i}at+\int_{-\infty}^{\infty}\left[\mathrm{e}^{\mathrm{i}tu}-1-\frac{\mathrm{i}tu}{1+u^2}\right]\cdot\frac{1+u^2}{u^2}\mathrm{d}G(u)\right\},$$

其中 i 为虚数单位，a 为任一实数，G 为非降左连续函数，$G(-\infty)=0$，$G(+\infty)<\infty$，被积函数当 $u=0$ 时应理解为当 $u\to0$ 时的极限. φ 与 (a,G) 是相互唯一确定的.

此定理给出了无穷可分分布族的构造. 例如：

推论 10.2.7 在定理 10.2.6 中，

(1) 令 $G(u)\equiv0$，则 $\varphi(t)=\mathrm{e}^{\mathrm{i}at}$，为退化到 a 的分布.

(2) 令

$$G(u)=\begin{cases}0, & 当\ u\leqslant 0,\\ \sigma^2, & 当\ u>0.\end{cases}$$

则 $\varphi(t)=\mathrm{e}^{\mathrm{i}at+\frac{\sigma^2t^2}{2}}$，为正态分布 $N(a,\sigma^2)$.

(3) 令 $a=\dfrac{\lambda}{2}$，

$$G(u)=\begin{cases}0, & 当\ u\leqslant 1,\\ \dfrac{\lambda}{2}, & 当\ u>1.\end{cases}$$

其中 $\lambda>0$，则 $\varphi(t)=\exp[\lambda(e^{it}-1)]$为泊松分布 $P(\lambda)$.

当一个无穷可分分布具有有限方差时，它的表达形式可变得简单一些.

定理 10.2.8 具有有限方差的分布 F 是无穷可分的，当且仅当它的特征函数 φ 可表为

$$\varphi(t)=\exp\{i\gamma t+\int_{-\infty}^{\infty}[e^{itu}-1-itu]\frac{1}{u^2}dK(u)\},$$

其中 i 为虚数单位，γ 为有限实数，K 为非降左连续函数，$K(-\infty)=0$，$K(+\infty)<\infty$.

定理 10.2.9 设独立随机变量组列 $\{X_{nk},1\leqslant k\leqslant k_n\}$，$n=1,2,\cdots$满足定理 10.2.4 的条件，$S_n=\sum_{k=1}^{k_n}X_{nk}$. 则下列两命题等价：

(1) $S_n\xrightarrow{d}F$，F 的特征函数由定理 10.2.6 中的(a,G)刻画.

(2) $$\lim_{n\to\infty}\sum_{k=1}^{k_n}\left[a_{nk}(\tau)+\int_{-\infty}^{\infty}\frac{x}{1+x^2}dF_{nk}(x+a_{nk}(\tau))\right]=a,\quad\forall\tau>0,$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{-\infty}^{u}\frac{x^2}{1+x^2}dF_{nk}(x+a_{nk}(\tau))=G(u),\quad\forall\tau>0,\text{及 }u\in C(G),$$

其中 F_{nk}为 X_{nk}的分布函数，$a_{nk}(\tau)=\int_{[|x|<\tau]}x dF_{nk}(x)$，$C(G)$为 G 的连续点集.

推论 10.2.10 设$\{X_{nk},k=1,\cdots,k_n\}$，$n=1,2,\cdots$，满足定理 10.2.4 的条件，$S_n=\sum_{k=1}^{k_n}X_{nk}$，$F_{nk}$为 X_{nk}的分布函数，则

(1) $S_n\xrightarrow{d}0$，当且仅当

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{[|x|\geqslant y]}dF_{nk}(x)=0,\quad\text{对某个 }y>0;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\left\{\int_{[|x|<y]}x^2\mathrm{d}F_{nk}(x)-\left[\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)\right]^2\right\}=0,\quad \forall\ y>0;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)=0,\quad \text{对某个 } y>0.$$

(2) $S_n\xrightarrow{\mathrm{d}}P(\lambda)$，当且仅当

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}F_{nk}(x)=0,\quad \forall\ x<0;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}(1-F_{nk}(x))=\begin{cases}\lambda,\text{当 } 0<x<1,\\ 0,\text{当 } 1<x<\infty;\end{cases}$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\left\{\int_{[|x|<y]}x^2\mathrm{d}F_{nk}(x)-\left[\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)\right]^2\right\}=0,\quad 0<y<1;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)=0,\quad \text{当 } 0<y<1.$$

(3) $S_n\xrightarrow{\mathrm{d}}N(0,1)$，当且仅当

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{[|x|\geqslant y]}\mathrm{d}F_{nk}(x)=0,\quad \forall\ y>0;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\left\{\int_{[|x|<y]}x^2\mathrm{d}F_{nk}(x)-\left[\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)\right]^2\right\}=1,\quad \forall\ y>0;$$

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{[|x|<y]}x\mathrm{d}F_{nk}(x)=0,\quad \forall\ y>0.$$

10.2.2 独立随机变量列的部分和

设 $\{X_n,n\geqslant 1\}$ 为独立随机变量列，记 $S_n=\sum_{i=1}^{n}X_i,n=1,2,\cdots$. 对常数列 $\{A_n\}$ 和 $\{B_n>0\}$，考虑 $S_n^*=(S_n-A_n)/B_n$ 的极限分

布. 记 $A_0=0, a_n=A_n-A_{n-1}, n=1,2,\cdots$，令 $\tilde{X}_{nk}=(X_k-a_k)/B_n$，$1\leqslant k\leqslant n, n=1,2,\cdots$，则 $S_n^*=\sum_{k=1}^{n}\tilde{X}_{nk}$. 因此，这是普遍中心问题的一个特例.

定理 10.2.11 设 $A_n, B_n, \tilde{X}_{nk}$ 与 S_n^* 如上所定义，S_n^* 的极限分布存在且非退化，则 $\{\tilde{X}_{nk}\}$ 满足一致渐近可忽略条件(10.2.1)，当且仅当 $B_n\to\infty$，且 $B_n/B_{n+1}\to 1$，当 $n\to\infty$.

显然，在满足一致渐近可忽略条件的前提下，S_n^* 可能的极限分布为无穷可分分布族的一个子族.

定义 10.2.12 设特征函数 φ 满足，对 $\forall\ c\in(0,1)$，$\dfrac{\varphi(t)}{\varphi(ct)}$ 在 $t\in\mathbf{R}^1$ 上定义了一个特征函数，则称 φ 是**自分解的**(self-decomposible)，相应的分布函数称为**自分解分布函数**.

定理 10.2.13 设 $\{X_n, n\geqslant 1\}$ 为独立随机变量列，$\{a_n\}$，$\{B_n>0\}$ 分别为常数列，$\tilde{X}_{nk}=(X_k-a_k)/n$，$1\leqslant k\leqslant n$，$n\geqslant 1$. 若 $\{X_{nk}, 1\leqslant k\leqslant n, n\geqslant 1\}$ 满足一致渐近可忽略条件(10.2.1)，则 $S_n^*=\sum_{k=1}^{n}\tilde{X}_{nk}$ 的一切可能的极限分布族为自分解分布族.

自分解分布族是无穷可分分布族的真子集.

定理 10.2.14 自分解特征函数 φ 对应着定理 10.2.6 中的 (a,G) 满足下述条件：

(1) G 在 $\mathbf{R}^1$-$\{0\}$ 上存在着左，右导数，分别记为 g_L 和 g_R.

(2) $\dfrac{1+x^2}{x}g_L(x)$ 和 $\dfrac{1+x^2}{x}g_R(x)$ 为非增函数.

推论 10.2.15 退化分布和正态分布均为自分解分布，但泊松分布不是自分解分布.

10.2.3 独立同分布随机变量列的部分和

定义 10.2.16 若 F 的特征函数 φ 满足：对任一对实数 a_1, a_2

和任一对正数 b_1, b_2，存在实数 a 和正数 b，使得

$$e^{ita_1}\varphi(b_1 t) \cdot e^{ita_2}\varphi(b_2 t) = e^{ita}\varphi(bt).$$

则分布函数 F 称为是**稳定的**(stable).

定理 10.2.17 设 $\{X_n, n \geqslant 1\}$ 为独立同分布随机变量列，$S_n = \sum_{i=1}^{n} X_i, n = 1,2,\cdots$，对任意常数列 $\{A_n\}$ 和 $\{B_n > 0\}$，$(S_n - A_n)/B_n$ 的一切可能的极限分布族是稳定族.

定理 10.2.18 分布函数 F 为稳定的，当且仅当它的特征函数有形式

$$\varphi(t) = \exp\{iat - c|t|^{\alpha}[1 + i\beta \cdot \operatorname{sgn}(t) \cdot w(t,\alpha)]\},$$

其中 a 为任意常数，$0 < \alpha \leqslant 2$，$|\beta| \leqslant 1$，$c \geqslant 0$，$\operatorname{sgn}(t)$ 为 t 的符号，

$$w(t,\alpha) = \begin{cases} \operatorname{tg}\left(\dfrac{\pi\alpha}{2}\right), & \text{当 } \alpha \neq 1, \\ \dfrac{\alpha}{\pi}\log|t|, & \text{当 } \alpha = 1. \end{cases}$$

其中 α 称为**特征指数**(characteristic exponent)，c 称为**恰当参数**(proper parameter).

推论 10.2.19 在稳定分布特征函数 φ 的表示式中，

(1) 当恰当参数 $c=0$ 时，为退化分布.

(2) 当 $c>0$，特征指数 $\alpha=2$ 时，为正态分布 $N(a,\sigma^2)$，其中 $\sigma^2=2c$.

(3) 当 $c>0$，$0<\alpha<2$ 时，p 阶矩存在当且仅当 $p<\alpha$.

11 点估计

11.1 统计模型

11.1.1 总体与样本

数理统计学(mathematical statistics)是研究搜集和分析随机数据的科学方法的一门学科,是应用数学的一个重要分支.

定义 11.1.1 在数理统计学中,研究对象的全体所构成的集合称为**总体**(population).

总体是一个非常广泛的概念.在许多实际问题中,总体是一个包含实在对象的具体概念,如一群人或一批产品等,而在另一些问题中,总体又可能是一个抽象的概念.

总体中的**元素**(elements)或**个体**(individuals)的个数可能是有限的或无限的.在前一场合,我们称之为有限总体,而后者称为无限总体.有时,一个有限总体中的个体数目非常大,可近似看成是无限总体.

数理统计学所关心的不是总体本身,而是总体的某些数量的性质.例如,在一群人所构成的总体中,我们关心的是这群人的某一生理指标(例如身高、体重或血压等等)的分布情况;或同时考虑多个指标的分布情况.

定义 11.1.2 假设总体的某一个或多个指标变量的取值情况已知,我们就可以计算出它(或它们)的分布函数,而这分布函数在总体上确定了一个概率分布,称为**总体分布**(distribution of population).

实际上总体分布总是全部或部分地未知的.而数理统计学的

全部任务就在于设法全部或部分地了解总体分布的有关信息.

假如某个总体(例如一群人)的所有个体的指标变量值(例如身高)都已测量好了,则总体分布就唯一地确定了.从数理统计的角度来看,对这个总体的关于这个指标的分布的认识已经达到了尽善尽美的状态.事实上数理统计学所面临的任务永远是:已经知道了总体的一部分个体的指标变量值,以此为出发点来推断总体分布的性质.

定义 11.1.3 总体的一部分个体的指标变量值称为一个**样本**(sample).

数理统计学的任务就是用一定的数学方法,根据已知的样本去推断总体的未知性质.为此,要求有获取样本的科学方法.

定义 11.1.4 **随机抽样**(random sampling)是指这样一种获取样本的方法:每一可能的样本被抽取的机会是可计算的.

上述定义中有两点需要说明.第一,所谓"被抽取的机会是可计算的"是指按照某种概率结构来抽取样本,以便在使用样本去进行推断时,能够使用概率论的理论.因为绝大部分的数理统计方法是建立在概率论的理论之上的.第二,"随机"抽样并不意味着每一可能的样本有同等的机会被抽取.

以下的各章节中,凡提到抽样,都是指随机抽样.

定义 11.1.5 对有限总体来说,有两种基本的抽样方法:**有放回抽样**(sampling with replacement)——是指在抽样过程中依次抽取的每一个体,经过记录指标变量值之后,仍放回到原总体之中;**无放回抽样**(sampling without replacement)——是指在抽样过程中依次抽取的每一个体,经过记录其指标变量值之后,不再放回到总体中去.

有放回抽样与无放回抽样的一个重要区别在于:在有放回抽样过程中,总体始终保持不变;而在无放回抽样过程中,总体在不断缩小.

但是当总体是无限的，或有限总体中元素（个体）特别多，而抽取的样本相对来说较小时，无放回抽样可近似地看作是有放回抽样，因为通常有放回抽样产生较简单的样本概率结构.

每一抽样过程蕴含着产生一个可能的样本，由于这样本是未被实现的，因此是一个随机变量（或随机向量，随机矩阵）. 但当抽样过程被实施以后，我们便得到一组真实的数据，称为**样本值**（sample value）.

在数理统计的理论研究中，对某一统计方法的性状分析和优劣的评价是根据这一统计方法在多次使用过程中的"总的"或"平均"的效果来衡量的，因此，样本被看成是随机变量（或随机向量，随机矩阵）更为自然. 对某个实际的统计问题，一个统计方法的使用是一次性的，我们面临着的是一组具体的样本值.

对于总体和样本，我们规定用 X 或其它大写拉丁字母记总体，实际上是总体的一个或一组指标值，假定有 d 个指标值，则称总体是 d 维的. 用（$X_1, X_2, \cdots, X_n$）或带下标的大写拉丁字母表示样本，其中每个 X_i 代表一个个体的相应指标变量值，都是 d 维的，我们称样本是 d 维的（d-dimensinnal sample）. n 为样本中所包含个体的个数，称为样本的大小（size）或容量（content）. 我们用（$x_1, x_2, \cdots, x_n$）或其它带下标的小写拉丁字母表示一组（已实现的）样本值，其维数和大小同上定义.

定义 11.1.6　简单样本（simple sample）是指这样的样本（$X_1, X_2, \cdots, X_n$），它的分量 $X_i, i=1, \cdots, n$ 是独立同分布的随机变量（向量）.

以下章节大部分是讨论一维简单样本，简称为样本.

对有限总体采用有放回抽样，假定每次抽取（一个个体）时，总体中的每个个体都有"同等"的机会被抽到，则样本的每个分量与总体有相同的分布，且分量之间是相互独立的，这样就产生一个简单样本. 此外，还有一种不存在真实总体的简单样本. 设做某种科

学实验，在各种条件都不改变的前提下，互不干扰地重复 n 次试验，每一次试验产生一个(或一组)试验数据，这样也产生一个简单样本；每一次试验的结果构成样本的一个分量.在理论上，可以虚构一个由所有"可能的"试验结果所构成的，具有一定分布的总体，而每次试验相当于从这虚构的总体中进行一次有放回抽样.

11.1.2 统计推断与统计量

总体分布总是全部或部分地未知的.**统计推断**(statistical inference)的目的就是要根据样本所提供的信息去推断总体分布中的未知因素.根据所要解决的任务的不同就有各种不同的统计推断的问题.下面介绍一些常见的统计推断问题.

定义 11.1.7 设总体分布为 F，与 F 有关的一些参数是未知的，要估计这些参数的真值，这样的统计推断问题称为**估计**(estimation).根据提供答案的方式不同，又分**点估计**(point estimation)和**区间估计**(interval estimation)两大类.前者是要给出关于参数真值的一个点；后者是要给出参数真值可能存在的一个区间.根据对总体分布已经掌握的知识的程度的不同，又分**参数估计**(parametric estimation)和**非参数估计**(nonparametric estimation).参数估计是指当总体分布由有限个参数唯一决定时，对这些参数的估计.非参数估计是指当总体分布 F 不能由有限个参数唯一决定时，要估计它的某些重要的数字特征，例如均值、方差等等.

定义 11.1.8 对总体分布 F 作一假设，然后以样本为依据制定一个方案，当样本取值于某一范围时便肯定这个假设，否则便否定之，这样的推断问题称为**假设检验**(test of hypothesis).

定义 11.1.9 假定变量 y 与其它一些变量 $x_1,x_2,\cdots,x_p$ 之间有着受随机误差影响的函数关系：

$$y = f(x_1,x_2,\cdots,x_p) + e.$$

在上式中 f 为未知或部分未知的函数，e 表示随机误差. 我们的目的是要找到函数 f(通常是估计决定 f 的一些未知参数). 这样的统计推断问题称为**回归分析**(regression analysis).

对同一统计推断问题，可能有不同的统计方法，这些方法根据对总体已知程度的不同，理论准则的不同以及实际应用的不同，各有优缺点. 了解这些内容对实际应用中选择哪种统计方法的决定有重要的指导意义.

各种统计方法归根结底是依据样本. 样本中包含了总体分布的信息，通过对样本进行一定的加工，提取这些信息并集中起来用于推断总体的信息.

定义 11.1.10 设$(X_1,\cdots,X_n)$为一个样本，它的一个与总体分布无关的函数(或向量函数)$f(X_1,\cdots,X_n)$称为一个**统计量**(statistics).

统计量必须与总体分布无关，更确切地说，统计量必须不依赖任何未知的量. 另一方面，一个“好的”统计量必须能够充分“压缩”总体分布中的有关信息，才能在统计推断中提高效率.

例如，某一总体的分布为 $N(a,1)$，a 为未知参数. $X_1,\cdots,X_n$ 为从该总体中得到的一个简单样本. 记 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，则由正态分布的性质易知 $\overline{X}$ 服从 $N\left(a,\frac{1}{n}\right)$ 分布，且 $\overline{X}$ 的取值以 99.97%的概率落在区间$\left(a-\frac{3}{\sqrt{n}},a+\frac{3}{\sqrt{n}}\right)$之内. 因此当 n 较大时，$\overline{X}$ 取值离 a 较远的可能性就很小. 因此用 $\overline{X}$ 去近似 a 是适宜的.

由上例可见，判断一个统计量是否成为一个高效率的推断手段，前提是要精确地或尽可能近似地掌握统计量的分布. 第 6 章中所介绍的常用分布绝大部分在统计推断中有用，其中最重要的是正态及由正态派生出来的分布，如 χ^2 分布，t 分布及 F 分布等.

在讨论样本或统计量的分布时，常考虑的不是分布函数，而是

密度或概率值.

定义 11.1.11 对一个随机变量 X 及其分布,它的**频率函数**(frquency function)$f(x)$为

(1) 当 X 为连续型变量时,它的密度函数.

(2) 当 X 为离散型变量时,它在 x 点的概率值.

11.2 无偏估计

11.2.1 无偏估计的定义

设总体分布由有限个未知参数 $\theta=(\theta_1,\cdots,\theta_p)$所决定,记为 F_θ. θ 可能取值的范围称为**参数空间**(parameter space),记作 Θ. 设 $X_1,\cdots,X_n$ 为自该总体中抽出的一个样本.

定义 11.2.1 若用一个 p 维统计量 $\hat{\theta}=(\hat{\theta}_1(X_1,\cdots,X_n),\cdots,\hat{\theta}_p(X_1,\cdots,X_n))$去估计 θ 的真值,则称 $\hat{\theta}$ 为 θ 的一个**点估计**(point estimate).

对一个估计的基本要求是希望它尽量不要偏离 θ 的真值. 但由于估计作为样本的函数也是一个随机变量,上述要求只能在下述的两个意义上来实现. 首先,我们希望在多次使用这个估计时,"平均"来看,估计的 θ 是正确的,换句话说,估计的正负偏差应能相互抵消. 其次,我们希望,在重复使用一个估计时,其"平均累计误差"应尽可能小. 由第一个要求,我们给出

定义 11.2.2 设 $\hat{\theta}$ 为未知参数的一个估计,满足

$$E\hat{\theta}=\theta,$$

则称 $\hat{\theta}$ 为 θ 的一个**无偏估计**(unbiased estimate).

在上述定义中,当 $\theta=(\theta_1,\cdots,\theta_p)$为高维的,则 $E\hat{\theta}=(E\hat{\theta}_1,\cdots,E\hat{\theta}_p)$,其中$(\hat{\theta}_1,\cdots,\hat{\theta}_p)=\hat{\theta}$.

由大数定律,当反复用 $\hat{\theta}$ 去估计参数真值 θ 时,$\hat{\theta}$ 的平均值便

收敛到真值 θ.

下面给出两个常用的无偏估计.

例 11.2.3 设总体均值为 a，方差为 σ^2，$X_1,\cdots,X_n$ 为简单样本. 定义 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$. 则 $\overline{X}$ 为 a 的无偏估计，S^2 为 σ^2 的无偏估计，分别称为**样本均值**(sample mean)和**样本方差**(sample variance).

衡量一个估计对于参数真值的"平均累积误差"，在统计学的理论和应用中常用下述的估计量

定义 11.2.4 设 θ 是一个一维未知参数(或参数的某个分量)，$\hat{\theta}$ 是 θ 的一个估计，则定义 $\hat{\theta}$ 的**均方误差**(mean square error)为 $\mathrm{MSE}(\hat{\theta})=E(\hat{\theta}-\theta)^2$.

另一个可用的衡量估计的平均累积误差的量是

定义 11.2.5 设 θ 与 $\hat{\theta}$ 分别为 1 维参数与它的估计. 则 $\hat{\theta}$ 的**平均绝对误差**(mean absolute error)定义为 $\mathrm{MAE}(\hat{\theta})=E|\hat{\theta}-\theta|$.

均方误差与平均绝对误差作为估计误差的两种不同度量各有其优缺点. 前者具有数学上易于处理的优点而得到了广泛的应用，而后者具有稳健性(robustness)的优点，在现代统计理论中日益得到重视(见第 14 章).

若 $\hat{\theta}$ 是 θ 的一个无偏估计，则 $\mathrm{MSE}(\hat{\theta})=\mathrm{var}(\hat{\theta})$. 因此对无偏估计来说，方差是衡量其平均累积误差大小的重要的量.

传统的统计理论对估计的无偏性非常重视. 但过分强调这一点就会产生一些困难. 第一，在某些估计问题中无偏估计并不存在；第二，有些无偏估计即使存在，其表示形式也过于繁复，不合实用. 第三，无偏估计的均方误差不一定能达到最小.

11.2.2 最小方差无偏估计与充分完备统计量

对同一总体分布的参数，可以构造许多不同的无偏估计. 例如

对总体均值 a，任意构造样本 $X_1,\cdots,X_n$ 的一个函数 $T=\sum_{i=1}^{n}c_iX_i$，其中 $c_1,\cdots,c_n$ 为常数，$\sum_{i=1}^{n}c_i=1$．则显然具有这种形式的 T 都是 a 的无偏估计．如前所述，方差是衡量一个无偏估计的"平均累积误差"大小的量．因此，通常希望找到方差尽可能小的估计．

定义 11.2.6 设 θ 为总体的一维参数，$\hat{\theta}$ 为 θ 的一个无偏估计．若对 θ 的任一无偏估计 $\tilde{\theta}$ 及 θ 的任一可能值，有

$$\mathrm{var}(\hat{\theta})\leqslant\mathrm{var}(\tilde{\theta}).$$

则称 $\hat{\theta}$ 为 θ 的**最小方差无偏估计**（minimum variance unbiased estimate）．

理论上，寻找最小方差无偏估计要依赖于充分完备统计量．

定义 11.2.7 设 $X_1,X_2,\cdots,X_n$ 为取自总体分布 F_θ 的样本，θ 为（有限维的）未知参数．T 为一个（有限维的）统计量．若当给定 T 时，样本 $X_1,\cdots,X_n$ 的条件分布与未知参数 θ 无关，则称 T 为**充分统计量**（sufficient statistic）．

充分统计量的概念是由英国著名统计学家费希尔（Fisher）提出的，其含义是：样本中包含关于未知参数 θ 的信息全部压缩在充分统计量之中了．而这一点是通过比较样本的无条件分布和给定 T 之后的条件分布看出来的．

定义 11.2.8 设 $X_1,\cdots,X_n$ 为取自总体分布 F_θ 的一个样本，$\theta\in\Theta$ 为未知参数，T 为一个统计量．若对 T 的任一函数 $g(T)$ 满足

$$Eg(T)=0,\quad \text{任意 } \theta\in\Theta,$$

就有 $g(T)=0$，则称 T 为**完备统计量**（complete statistic）．

若一个统计量 T 既是充分的又是完备的，则称它为充分完备统计量．利用充分完备统计量便可得到关于最小方差无偏估计的一个重要结论．

定理 11.2.9 莱赫曼-谢菲(Lehman-Scheffe)定理 设 $\theta=(\theta_1,\cdots,\theta_p)$ 为总体分布 F_θ 的一个 p 维未知参数,$\theta\in\Theta$. $X_1,\cdots,X_n$ 是取自该总体的一个样本,$T=T(X_1,\cdots,X_n)$ 为一充分完备统计量(一般也是 p 维的). 对 θ 的任一函数 $g(\theta)$,若存在一个仅依赖于 T 的无偏估计 $\hat{g}(T)$,则 $\hat{g}(T)$ 为 $g(\theta)$ 的最小方差无偏估计.

定理 11.2.9 给出了一个寻找最小方差无偏估计的方法,即:先找充分完备统计量 T,再找一个仅依赖于 T 的无偏估计,则这无偏估计即为所求. 下面一个定理给出寻找充分统计量的简单方法.

定理 11.2.10 设样本 $X_1,\cdots,X_n$ 取自总体分布 F_θ,有频率函数

$$f(x_1,\cdots,x_n|\theta) = g(T(x_1,\cdots,x_n),\theta)\cdot h(x_1,\cdots,x_n),$$

其中 $T(x_1,\cdots,x_n)=(T_1(x_1,\cdots,x_n),\cdots,T_p(x_1,\cdots,x_n))$,与参数 θ 无关,$h(x_1,\cdots,x_n)$ 与 θ 无关,则 $T(x_1,\cdots,x_n)$ 为充分统计量. 上述条件也是必要的.

关于充分完备统计量,要分两种情形讨论. 首先考虑总体 X 的取值区域与 θ 无关的情形. 例如正态分布 $N(a,\sigma^2)$,其中 a,σ^2 为未知参数,而总体 X 的取值区域为整个实数轴,与 a,σ^2 的值无关. 这时有下述的定理.

定理 11.2.11 设总体 X 的取值区域与未知参数 θ 无关,有频率函数

$$f(x,\theta) = c(\theta)h(x)\exp\left\{\sum_{i=1}^{p}a_i(\theta)T_i(x)\right\},$$

其中$(a_1(\theta),\cdots,a_p(\theta))$关于 θ 是 1-1 对应的(即对每个$(a_1(\theta),\cdots,a_p(\theta))$,有唯一的 θ 值与之对应),则对简单样本 $X_1,\cdots,X_n$ 有充分完备统计量

$$T(X_1,\cdots,X_n) = \left\{\sum_{i=1}^{n}T_1(X_i),\cdots,\sum_{i=1}^{n}T_p(X_i)\right\}.$$

例 11.2.12 设总体分布 $N(a,\sigma^2)$，$\theta=(a,\sigma^2)$ 为二维未知参数. 其密度函数为

$$f(x,\theta)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{a^2}{2\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}x^2+\frac{a}{\sigma^2}x\right\}$$
$$=c(\theta)h(x)\exp\{a_1(\theta)T_1+a_2(\theta)T_2\},$$

这里 $h(x)\equiv 1$，$c(\theta)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{a^2}{2\sigma^2}}$，$a_1(\theta)=\frac{a}{\sigma^2}$，$a_2(\theta)=-\frac{1}{2\sigma^2}$，$T_1(x)=x$，$T_2(x)=x^2$. 因此对简单样本 $X_1,\cdots,X_n$，充分完备统计量为 $T=\left(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2\right)$. 而 $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$，$S^2=\frac{1}{n-1}\left[\sum_{i=1}^n X_i^2-\frac{1}{n}\left(\sum_{i=1}^n X_i\right)^2\right]$ 分别为 a 和 σ^2 的无偏估计(见 11.2.3)且均为 T 的函数. 故由定理 11.2.9 知 $\overline{X}$ 与 S^2 分别为 a 和 σ^2 的最小方差无偏估计.

另一种情形是总体 X 的取值范围与 θ 有关. 例如，总体分布为区间$[\theta_1,\theta_2]$上的均匀分布，这里 $\theta=(\theta_1,\theta_2)$ 为未知参数. 这时定理 11.2.11 不能使用. 但在常见的场合容易确定完备统计量.

定理 11.2.13 设 p 维统计量 $T=(T_1,T_2,\cdots,T_p)$ 有频率函数 $g(t_1,\cdots,t_p|\theta)$，在 p 维空间的一个连通区域 $C(\theta)$ 内大于零，其它地方等于零. 若 $C(\theta)$ 随 θ 连续变化，且对 p 维空间中任一点 $t=(t_1,\cdots,t_p)$，存在 θ_t 使得 $C(\theta_t)$ 收缩到一点 t，则 T 为完备统计量.

设某个统计量 T 满足定理 11.2.13，因而为完备统计量，再利用定理 11.2.10 可以验证它的充分性.

例 11.2.14 设 $X_1,\cdots,X_n$ 为简单样本，取自总体为(θ_1,θ_2)上的均匀分布，$-\infty<\theta_1\leqslant\theta_2<\infty$. 考虑二维统计量 $T=(X_{(1)},X_{(n)})$，其中 $X_{(1)}=\min\limits_{1\leqslant i\leqslant n}X_i$，$X_{(n)}=\max\limits_{1\leqslant i\leqslant n}X_i$. 则根据顺序统计量的分布理论(见第 6 章)知 T 有联合概率密度

$$g(t_1,t_2|\theta_1,\theta_2)=\frac{n(n-1)}{(\theta_2-\theta_1)^n}(t_2-t_1)^{n-2},\theta_1\leqslant t_1\leqslant t_2\leqslant\theta_2.$$

由定理 11.2.13 可知，$T=(X_{(1)},X_{(n)})$为完备统计量. 而样本联合密度为

$$f(x_1,\cdots,x_n|\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n}I_{\{\theta_1\leqslant x_{(1)}\leqslant x_{(n)}\leqslant\theta_2\}}.$$

由定理 11.2.10 可知，$T=(X_{(1)},X_{(n)})$也是充分统计量，因此是充分完备统计量. 容易算得

$$EX_{(1)}=\theta_1+\frac{\theta_2-\theta_1}{n+1},\ EX_{(n)}=\theta_2-\frac{\theta_2-\theta_1}{n+1}.$$

因此 θ_1 的最小方差无偏估计为$\dfrac{nX_{(1)}-X_{(n)}}{n-1}$，$\theta_2$ 的最小方差无偏估计为$\dfrac{nX_{(n)}-X_{(1)}}{n-1}$.

11.2.3 无偏估计的方差下界

当找到最小方差无偏估计时，无偏估计的方差下界便可以通过计算该估计的方差来得到. 但需要有一种方法、不依赖于最小方差无偏估计而直接计算无偏估计的方差下界.

定义 11.2.15 设样本$(X_1,\cdots,X_n)$的联合分布有频率函数 $f(x_1,\cdots,x_n|\theta)$，其中 θ 为未知参数. 该样本关于 θ 的**似然函数**(likelihood function)为

$$L(\theta)=f(X_1,\cdots,X_n|\theta).$$

事实上似然函数就是把样本直接代入到它的频率函数中所得到的，重点是把它看成为 θ 的函数，因此记作 $L(\theta)$.

定理 11.2.16 克拉美-饶(Cramer-Rao)定理　设 T 为 θ 的一个无偏估计，且似然函数 $L(\theta)$满足条件

(1) $E\left[\dfrac{\partial\log L(\theta)}{\partial\theta}\right]=0$,

(2) $E\left[T\,\dfrac{\partial\log L(\theta)}{\partial\theta}\right]=1$,

则

$$\mathrm{var}(T) \geqslant E^{-1}\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right].$$

其中 $E^{-1}\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right]$称为 C-R 下界.

推论 11.2.17 设总体分布有频率函数 $f(x|\theta)$, $X_1,\cdots,X_n$ 为简单样本,则 C-R 下界有形式

$$\frac{1}{E\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right]} = \frac{1}{nE\left[\left(\frac{\partial \log f(X_1|\theta)}{\partial \theta}\right)^2\right]}.$$

因此,当样本为独立同分布时,C-R 下界是$\frac{1}{n}$数量级的.

当 $\theta=(\theta_1,\cdots,\theta_p)$,且 $p>1$ 时,可对 θ 的各个分量使用定理 11.2.16 来得到无偏估计的 C-R 下界.

例 11.2.18 设总体分布为 $N(a,\sigma^2)$,其中 a,σ^2 为未知参数,$X_1,\cdots,X_n$ 为简单样本,则

$$E\left[\frac{\partial \log L(a,\sigma^2)}{\partial a}\right]^2 = \frac{1}{\sigma^4}E\left[\sum_{i=1}^{n}(X_i-a)^2\right] = \frac{n}{\sigma^2},$$

$$E\left[\frac{\partial \log L(a,\sigma^2)}{\partial \sigma^2}\right]^2 = E\left[\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\frac{(X_i-a)^2}{\sigma^2}-n\right)\right]^2 = \frac{n}{2\sigma^4}.$$

因此 a 的 C-R 下界为$\frac{\sigma^2}{n}$,而 σ^2 的 C-R 下界为$\frac{2\sigma^4}{n}$. 另一方面,由例 11.2.12 知,a 的最小方差无偏估计为 $\overline{X}$,其方差为$\frac{\sigma^2}{n}$,而 σ^2 的最小方差无偏估计为 S^2,其方差为$\frac{2\sigma^4}{n-1}$. 由此可见,C-R 下界有时失之过小.

在 C-R 定理(11.2.16)的条件下,有

推论 11.2.19 $E\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right]$,

其中 $-E\left[\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right]$ 称为**费希尔信息**(Fisher information).

11.3 极大似然估计与矩估计

11.3.1 极大似然估计

极大似然估计(maximum likelihood estimate)是英国著名统计学家费希尔首先提出的一种应用非常广泛的估计方法.设样本$(X_1,\cdots,X_n)$有频率函数$f(x_1,\cdots,x_n|\theta)$,其中θ为p维未知参数.由定义11.2.15,似然函数为

$$L(\theta)=f(x_1,\cdots,x_n|\theta).$$

当给定样本值及$\theta=\theta_0$时,$L(\theta_0)$表示样本取给定值的可能性的相对大小($L(\theta_0)$乘一个比例因子相当于概率值).费希尔认为应该在θ的取的所有可能取值中找这样的值$\hat{\theta}$去估计θ:样本取给定值的可能性当$\theta=\hat{\theta}$时达到最大,即$L(\hat{\theta})$应该达到最大,他把这一思想称为**似然原理**.

定义 11.3.1 设p维统计量$\hat{\theta}$满足:对θ的任一可能值θ_0有

$$L(\hat{\theta})\geqslant L(\theta_0),$$

则称$\hat{\theta}$为θ的**极大似然估计**.

例 11.3.2 设总体分布为$U(0,\theta)$:区间$[0,\theta]$上的均匀分布,其中$0<\theta$为未知参数,$X_1,\cdots,X_n$为简单样本.则似然函数为

$$L(\theta)=\theta^{-n}I_{(X_{(n)}\leqslant\theta)},$$

其中$X_{(n)}=\max\{X_i,1\leqslant i\leqslant n\}$.显然,极大似然估计为$\hat{\theta}=X_{(n)}$.

根据定义,求极大似然估计就是要求似然函数$L(\theta)$的极大值点.在很多场合下,似然函数$L(\theta)$关于变量θ有二阶偏导数存在.因此极大似然估计$\hat{\theta}$应该满足条件

$$\left.\frac{\partial L(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}=0,\left.\frac{\partial^2L(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}<0.\qquad(11.1)$$

因此,求出式(11.1)的全部解,其中之一就是极大似然估计.

当样本 $X_1,\cdots,X_n$ 独立同分布，有相同的频率函数 $f(x|\theta)$时，常求解与 11.1 的第一个方程等价的方程：

$$\frac{\partial \log L(\theta)}{\partial \theta}=\sum_{i=1}^{n}\frac{\partial \log f(X_i|\theta)}{\partial \theta}=0. \tag{11.2}$$

11.2 式称为**对数似然方程**(loglikelihood equation)，其中 $\log L(\theta)$称为**对数似然函数**.

例 11.3.3 设总体分布为 $N(a,\sigma_0^2)$，其中 a 为未知参数，σ_0^2 为已知常数，$X_1,\cdots,X_n$ 为简单样本，要求均值 a 的极大似然估计. 此处，似然函数为

$$L(a,\sigma_0^2)=[2\pi\sigma_0^2]^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{1}^{n}(X_i-a)^2\right\},$$

对数似然函数为

$$\log L=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma_0^2-\frac{1}{2\sigma_0^2}\sum_{1}^{n}(X_i-a)^2.$$

对数似然方程为

$$\frac{\partial \log L}{\partial a}=\frac{1}{\sigma_0^2}\sum_{1}^{n}(X_i-a)=0.$$

因此 a 的极大似然估计为 $\hat{a}=\overline{X}=\frac{1}{n}\sum_{1}^{n}X_i$.

例 11.3.4 设总体分布为 $N(a,\sigma^2)$，其中 a,σ^2 均为未知参数，$X_1,\cdots,X_n$ 为简单样本，求 a 和 σ^2 的极大似然估计. 这时对数似然方程有两个：

$$\frac{\partial \log L}{\partial a}=\frac{1}{\sigma^2}\sum_{1}^{n}(X_i-a)=0,$$

$$\frac{\partial \log L}{\partial \sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{1}^{n}(X_i-a)^2=0.$$

先求解第一个方程得 $\hat{a}=\overline{X}$，将 $\hat{a}$ 代入第二个方程得 $\hat{\sigma^2}=\frac{1}{n}\sum_{1}^{n}(X_i-\overline{X})^2=\frac{n-1}{n}S^2$. 因此 $\overline{X}$ 与$\frac{n-1}{n}S^2$ 为 a 与 σ^2 的极大

似然估计. 这里 σ^2 的估计不是无偏的(见例 11.2.3).

实际求解似然方程常会遇到两个问题. 首先是常常无法求出解的直接表达式;其次是解的不唯一性. 为解决第一个问题,可求数值解. 这样能得到似然方程的一个近似解(但无法判明其是否仅仅是一个局部极值点).

下面介绍用牛顿-拉夫森(Newton-Raphson,N-R)方法求似然方程数值解的步骤. 设 $\hat{\theta}$ 为对数似然方程(11.2)的解,$\hat{\theta}_1$ 为 θ 的任一估计,对似然方程作泰勒展开,有

$$0=\left(\frac{\partial \log L}{\partial \theta}\right)_{\hat{\theta}}=\left(\frac{\partial \log L}{\partial \theta}\right)_{\hat{\theta}_1}+(\hat{\theta}-\hat{\theta}_1)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta^*},$$

其中 θ^* 位于 $\hat{\theta}_1$ 与 $\hat{\theta}$ 之间. 由上式可解得

$$\hat{\theta}=\hat{\theta}_1-\left(\frac{\partial \log L}{\partial \theta}\right)_{\hat{\theta}_1}\Big/\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta^*}.$$

当 $\hat{\theta}_1$ 距 $\hat{\theta}$ 较近时,θ^* 也距 $\hat{\theta}$ 较近,这时可在上式中用 $\hat{\theta}_1$ 代替 θ^*,得

$$\hat{\theta}_2=\hat{\theta}_1-\left(\frac{\partial \log L}{\partial \theta}\right)_{\hat{\theta}_1}\Big/\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\hat{\theta}_1}. \tag{11.3}$$

在此式中得到的 $\hat{\theta}_2$ 将比 $\hat{\theta}_1$ 距 $\hat{\theta}$ 更近. 若将 $\hat{\theta}_2$ 代替 $\hat{\theta}_1$ 代入上式式中便可得 $\hat{\theta}_3$,如此迭代 k 次,则 $\hat{\theta}_k$ 将非常接近 $\hat{\theta}$. 但是利用 N-R 方法所求的迭代解可能不收敛,或收敛的点不是 $\log L$ 的极大值点. 这方面的深入分析,超出了本书的范围. 但经验和理论分析证明,当样本容量 n 较大时,适当地选取初始值,迭代一两次便可得到一个相当满意的解. 例如,可以考虑用 θ 的矩估计或其它相合估计来作初始值.

11.3 式计算起来仍不够简便,在实际计算中费希尔建议用

$$E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)=-E\left(\frac{\partial \log L}{\partial \theta}\right)^2$$

去代替 $\frac{\partial^2 \log L}{\partial \theta^2}$,得

$$\hat{\theta}_2 = \hat{\theta}_1 - \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_1} \Bigg/ E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta_1}. \qquad (11.4)$$

用 11.4 式进行迭代称为**费希尔得分法**(Fisher scoring method).费希尔得分法与 N-R 法比较,在迭代的最初几步,前者收敛较快;而当迭代步数较多时,后者将比前者收敛得快.

例 11.3.5 设总体分布为柯西分布,有密度

$$f(x|\theta) = \frac{1}{\pi[1 + (x-\theta)^2]}, \quad -\infty < x < \infty.$$

$(X_1, \cdots, X_n)$为简单样本,求 θ 的极大似然估计,可用费希尔得分法来求数值解.似然方程为

$$\frac{\partial \log L}{\partial \theta} = 2\sum_{i=1}^{n} \frac{X_i - \theta}{1 + (X_i - \theta)^2} = 0.$$

首先计算

$$E\left(\frac{\partial^2 \log L}{\partial \theta}\right) = n\left(\frac{\partial^2 \log f}{\partial \theta^2}\right) = \frac{n}{\pi}\int_{-\infty}^{\infty} \frac{2[(x-\theta)^2 - 1]}{[1 + (x-\theta)^2]^3} dx$$

$$= \frac{4n}{\pi}\int_0^{\infty} \frac{(x^2 - 1)}{(1 + x^2)^3} dx = -\frac{n}{2}.$$

于是由 11.4 得迭代方程

$$\hat{\theta}_2 = \hat{\theta}_1 + \frac{4}{n}\sum_{i=1}^{n} \frac{(X_i - \hat{\theta})}{[1 + (X_i - \hat{\theta})^2]}.$$

至于初始值,这里推荐用样本中位数.

11.3.2 矩方法

定义 11.3.6 设 $X_1, X_2, \cdots, X_n$ 为简单样本.

(1) 样本 k 阶原点矩为

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, k = 1, 2, \cdots.$$

当 $k=1$ 时,$a_1 = \overline{X}$ 为**样本均值**.

(2) 样本 k 阶中心矩为

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k, k = 2,3,\cdots.$$

当 $k=2$ 时，$m_2=\frac{n-1}{n}s^2$，s^2 为**样本方差**.

定义 11.3.7 设总体分布的未知参数 $\theta=h(\alpha_1,\cdots,\alpha_r,\mu_2,\cdots,\mu_s)$，其中 $\alpha_l,l=1,\cdots,r$ 为总体的 l 阶原点矩，$\mu_k,k=2,\cdots,s$ 为总体的 k 阶中心矩，则 θ 的**矩估计**(estimation by moments)为

$$\hat{\theta} = h(a_1,\cdots,a_r,m_2,\cdots,m_s).$$

其中 $a_1,\cdots,a_r$ 为**样本原点矩**，$m_2,\cdots,m_s$ 为**样本中心矩**.

矩估计方法可概括为简单的一句话：用相应的样本矩去估计总体矩.

例 11.3.8 设 $c_s=\mu_3/\mu_2^{\frac{3}{2}}$，$c_k=\mu_4/\mu_2^2$ 分别为总体的偏度系数和峰度系数，则它们的矩估计分别为 $\hat{c}_s=m_3/m_2^{3/2}$，$\hat{c}_k=m_4/m_2^2$.

定理 11.3.9

(1) a_k 为 α_k 的无偏估计，即 $E(a_k)=\alpha_k$，估计的方差为 $\frac{1}{n}(\alpha_{2k}-\alpha_k^2)$.

(2) m_k 不是 μ_k 的无偏估计，偏差为 $O\left(\frac{1}{n}\right)$，即

$$E(m_k) = \mu_k + O\left(\frac{1}{n}\right).$$

估计的均方差为

$$E(m_k - \mu_k)^2$$
$$=\frac{1}{n}(\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2) + O\left(\frac{1}{n^2}\right).$$

注 在定理 11.3.9 中，$O\left(\frac{1}{n}\right)$ 表示与 $\frac{1}{n}$ 同级的无穷小量，$O\left(\frac{1}{n^2}\right)$ 表示与 $\frac{1}{n^2}$ 同级的无穷小量.

11.4 估计的大样本性质

11.4.1 基本概念

极大似然估计与矩估计以及其它的一些估计都是从某种直观的统计思想提出来的估计方法．它们既不必是无偏的，也不必是方差(或均方误差)最小的．在许多情况下直接计算估计的均值和方差均比较困难．概率论的极限理论的发展给数理统计学提供了一个有力的工具，使得他们能够研究当样本容量很大时，估计量的渐近性质。也就是说，设 $X_1,\cdots,X_n$ 为样本，研究当 $n\to\infty$时，统计量的性质．这种理论称为大样本理论．

设 θ 为待估参数，用 $\hat{\theta}_n$ 记基于容量为 n 的样本所产生的估计量．

定义 11.4.1 若对任意给定小的正数 ϵ 和 δ，存在 N，当 $n>N$ 时有

$$P(|\hat{\theta}_n-\theta|\geqslant\varepsilon)\leqslant\delta,$$

则称 $\hat{\theta}_n$ 为 θ 的**相合估计**(consistent estimate)．

相合估计的含义是：只要样本容量足够大，则估计量对真参数的偏离超过指定限度($|\hat{\theta}_n-\theta|\geqslant\varepsilon$)的概率可任意接近于零．从概率论的角度来看，就是要 $\hat{\theta}_n$ 依概率收敛到真参数 θ．从大样本的角度看，相合性是对估计的一个最基本的要求．

定义 11.4.2 设 $\hat{\theta}_n$ 为 θ 的估计，满足

$$P(\lim_{n\to\infty}\hat{\theta}_n=\theta)=1,$$

则称 $\hat{\theta}_n$ 为 θ 的**强相合估计**(strong consistent estimate)．

从概率论的角度看，强相合性是要求估计 $\hat{\theta}_n$“几乎处处”收敛到 θ，强于相合性(只要求依概率收敛)．

要研究一个估计的性质，了解它的分布是必不可少的．但某些

估计(例如极大似然估计)的分布非常复杂,以致无法给出明显的表达式.利用概率论中的中心极限定理,我们希望能找到估计量的极限分布,而在大多数场合下,这种极限分布是正态分布.

定义 11.4.3 设 $\hat{\theta}_n$ 作为 θ 的估计满足

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{d}} N(0,\sigma^2),\text{当 } n \to \infty,$$

其中 $\xrightarrow{\mathrm{d}}$ 表示依分布收敛.则称 $\hat{\theta}_n$ 为 θ 的**渐近正态估计**(asymptotieally normal estimate). σ^2 称为 $\hat{\theta}_n$ 的**渐近方差**(asymptotic variance),它可能与 θ 有关.

容易证明,渐近正态估计一定是相合估计,但不一定是强相合估计.

对某个待估参数 θ,如果存在着渐近正态估计,这样的估计可能并不唯一.因此渐近方差的大小就可以作为比较这些估计优劣的一个准则.

定义 11.4.4 设 $\hat{\theta}_n$ 为待估参数 θ 的一个渐近正态估计,渐近方差为 σ^2.若对 θ 的任意渐近正态估计 $\tilde{\theta}_n$,渐近方差记为 $\tilde{\sigma}^2$,有

$$\sigma^2 \leqslant \tilde{\sigma}^2,$$

则称 $\hat{\theta}_n$ 为 θ 的**最优渐近正态估计**(the best asymptotically normal estimate).

11.4.2 极大似然估计的大样本性质

在较一般的条件下,极大似然估计就是强相合估计.下面给出一个容易验证的充分条件.

定理 11.4.5 设总体 X 有频率函数 $f(x|\theta)$,它是 θ 的连续函数,其中 θ 为一维参数.设 $X_1,\cdots,X_n$ 为简单样本,$L(\theta)$为似然函数.若

(1) $E_\theta[\log f(X_1|\theta)]$存在,其中 E_θ 表示在真参数为 θ 时的期望.

(2) θ 的取值范围有界，或 $\lim\limits_{|\theta|\to\infty}\log f(x|\theta)=-\infty$.

(3) 对 θ 的任一可能值，总体分布为非退化.

则极大似然估计必为强相合估计.

在较一般的条件下，极大似然估计也是最优渐近正态估计.

定理 11.4.6 设总体 X 有频率函数 $f(x|\theta)$，其中 θ 为一维参数，设 $X_1,\cdots,X_n$ 为简单样本，$L(\theta)$ 为似然函数. 若

(1) $E\left[\dfrac{\partial\log f(X_1|\theta)}{\partial\theta}\right]=0$.

(2) $-E\left[\dfrac{\partial^2\log L}{\partial\theta^2}\right]=E\left[\left(\dfrac{\partial\log L}{\partial\theta}\right)^2\right]>0$.

则极大似然估计为 θ 的最优渐近正态估计. 即

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{\mathrm{d}}N(0,I(\theta)),$$

其中 $I(\theta)=E\left[\left(\dfrac{\partial\log f(X|\theta)}{\partial\theta}\right)^2\right]$.

在定理 11.4.6 中条件(2)相当于

$$\begin{aligned}0&=\frac{\partial}{\partial\theta}\int\cdots\int\frac{\partial\log L}{\partial\theta}L\mathrm{d}x_1\cdots\mathrm{d}x_n \qquad (11.5)\\&=\int\cdots\int\frac{\partial}{\partial\theta}\left[\frac{\partial\log L}{\partial\theta}L\right]\mathrm{d}x_1\cdots\mathrm{d}x_n\\&=E\left[\frac{\partial^2\log L}{\partial\theta}\right]+E\left[\left(\frac{\partial\log L}{\partial\theta}\right)^2\right].\end{aligned}$$

当样本容量 n 较大时，用定理 11.4.6 便可得到极大似然估计的近似分布.

11.4.3 矩估计的大样本性质

下面的定理表明矩估计一般总是强相合的.

定理 11.4.7 设未知参数 $\theta=h(\alpha_1,\cdots,\alpha_l,\mu_2,\cdots,\mu_k)$，其中 $\alpha_1,\cdots,\alpha_l$ 为总体原点矩，$\mu_2,\cdots,\mu_k$ 为总体中心矩，h 为连续函数，则 θ 的矩估计 $\hat{\theta}=h(a_1,\cdots,a_l,m_2,\cdots,m_k)$（见定义 11.3.7）为 θ 的强

相合估计.

一般设参数 θ 可表为总体原点矩和中心矩的函数(见定义11.3.11),但由于中心矩又可表为原点矩的函数,因此 θ 必可表为总体原点矩的函数.(例如,$\theta=\mu_2$:总体方差,则 $\theta=\alpha_2-(\alpha_1)^2$.)所以我们总假定

$$\theta = g(\alpha_1,\cdots,\alpha_k). \tag{11.6}$$

定理 11.4.8 设 α_k 为总体 k 阶原点矩,a_k 为样本 k 阶原点矩,则

(1) a_k 为 α_k 的无偏估计,即 $E(a_k)=\alpha_k$.

(2) $\mathrm{var}(a_k)=\dfrac{1}{n}(\alpha_{2k}-\alpha_k^2)$.

(3) $\mathrm{cov}(a_k,a_l)=\dfrac{1}{n}(\alpha_{k+l}-\alpha_k\alpha_l)$.

定理 11.4.9 设未知参数 θ 满足关系式(11.6),且总体分布有 $2k$ 阶矩存在,函数 $g(\alpha_1,\cdots,\alpha_k)$ 对各变元有连续的一阶偏导数,则 θ 的矩估计 $\hat{\theta}=g(a_1,\cdots,a_k)$ 是 θ 的渐近正态估计,即

$$\sqrt{n}\,(\hat{\theta}-\theta)\xrightarrow{\mathrm{L}} N(0,v(\theta)).$$

其中 $v(\theta)=\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{k}(\alpha_{i+j}-\alpha_i\alpha_j)\dfrac{\partial g}{\partial\alpha_i}\dfrac{\partial g}{\partial\alpha_j}$,为渐近方差.

例 11.4.10 设 c_s 和 c_k 分别为总体的偏度系数与峰度系数,$\hat{c}_s$ 和 $\hat{c}_k$ 分别为它们的矩估计,则 $\hat{c}_s$ 和 $\hat{c}_k$ 分别为 c_s 和 c_k 的渐近正态估计,其渐近方差分别为

$$(4\mu_2^2\mu_6-12\mu_2\mu_3\mu_5-24\mu_2^3\mu_4+9\mu_3^2\mu_4+35\mu_2^2\mu_3^2 \\ +36\mu_2^5)/(4\mu_2^5)$$

和

$$(\mu_2^2\mu_8-4\mu_2\mu_4\mu_5-8\mu_2^2\mu_3\mu_5+4\mu_4^3-\mu_2^2\mu_4^2+16\mu_2\mu_3^2\mu_4 \\ +16\mu_2^3\mu_3^2)/\mu_2^6.$$

当总体为正态时,其值分别为 6 和 24.

11.5 贝叶斯估计

11.5.1 基本概念

在前几节中，所讨论的估计方法与概念有一个重要的前提，即待估参数是一个确定的常量. 但是在某些实际问题当中，待估参数并不是一个常量，而是具有随机性的量. 例如，某工厂生产的产品的日废品率，显然是逐日波动的，在一段时期（例如一个月）中，将日废品率看成是一个常数就不符合实际情况. 因此，检验员每天进行抽样检验废品率时，实际上是检验当日的废品率. 假定该工厂过去一段时间内日废品率都较低，而今天抽样的结果中废品较多，这就有两种可能性：第一种可能性是废品率确实增高了；第二种可能性是废品率并无明显增高，而是碰巧抽到一个不好的样本. 因此应分别考虑.

由上面的例子可以得出一种不同的统计推断的观点，即贝叶斯（Bayes）学派的观点. 下面来介绍贝叶斯学派的一些基本概念. 首先，待估参数 θ 不再能看成为常数，而是随时间（或空间）推移而不断变化的随机变量.

定义 11.5.1 未知参数 θ 作为随机变量，它的分布 $H(\theta)$ 称为**先验分布**（prior distribution）.

先验分布可能是已知的，未知的或部分未知的. 我们首先假定先验分布是已知的，则它表示了我们在未取到当前的样本以前对 θ 的已有的知识. 当取得当前的样本以后，样本中包含着关于 θ 当前的信息. 显然，在估计 θ 的值时，最合理的作法是把关于 θ 的已有的知识（先验分布）和包含 θ 当前信息的知识（样本）综合起来，以此为出发点对 θ 进行推断. 具体的作法如下. 设先验分布函数为 $H(\theta)$，总体分布函数为 $F(x|\theta)$ 可看成是给定 θ 时的条件分布. 于

是对于简单样本 $X_1,\cdots,X_n$，$(X_1,\cdots,X_n,\theta)$ 的联合分布函数为 $F(x_1|\theta)\cdots F(x_n|\theta)H(\theta)$. 设样本 $X_1=x_1,\cdots,X_n=x_n$，于是 θ 在给定样本下的条件分布 $H(\theta|x_1,\cdots,x_n)$ 便可由上述联合分布计算出来.

定义 11.5.2 给定样本 $X_1=x_1,\cdots,X_n=x_n$ 时，θ 的条件分布 $H(\theta|x_1,\cdots,x_n)$ 称为 θ 的**后验分布**(postrior distribution).

后验分布是由先验分布和样本所唯一决定的；换言之，在后验分布中包含了我们对 θ 的过去和现在的全部知识，因此是我们对 θ 进行统计推断的唯一出发点.

如何从后验分布 $H(\theta|x_1,\cdots,x_n)$ 出发对 θ 进行估计呢？首先要给出一个评价估计量的准则.

定义 11.5.3 设 $L(\theta,a)$ 为一个非负函数，它表示对参数真值 θ，用 a 去估计时所产生的损失，称为**损失函数**(loss function).

现假定 $\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)$ 为当样本 $X_1=x_1,\cdots,X_n=x_n$ 时，所产生的对 θ 的一个估计，则用 $\hat{\theta}$ 去估计 θ 时，所造成的损失为 $L(\theta,\hat{\theta})$. 由于 θ 是随机变量. 衡量 $\hat{\theta}$ 的好坏就不能只考虑对某一个 θ 值的估计所造成的损失，而应考虑对 θ 的所有可能值的"平均损失".

定义 11.5.4 设样本 $X_1=x_1,\cdots,X_n=x_n$，$\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)$ 则用 $\hat{\theta}$ 去估计 θ 的**后验风险**(postrior risk)为

$$\begin{aligned}R_{\hat{\theta}}(x_1,\cdots,x_n) &= E[L(\theta,\hat{\theta})|x_1,\cdots,x_n]\\ &= \int L(\theta,\hat{\theta})\mathrm{d}H(\theta|x_1,\cdots,x_n).\end{aligned}$$

后验风险反映了由一次取样产生的估计的平均损失. 在一次性决策问题中，用后验风险来评价估计的优劣是适宜的. 但在多次性决策问题中，我们要求计算对样本 $X_1,\cdots,X_n$ 的取得各种可能值所造成的平均损失.

定义 11.5.5 对样本 $X_1,\cdots,X_n$，估计量 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 的**风险**(risk)为

$$R_{\hat\theta} = E[L(\theta,\hat\theta(X_1,\cdots,X_n)].$$

注 后验风险是把 $X_1=x_1,\cdots,X_n=x_n$ 固定下来，对 $L(\theta,\hat\theta)$ 关于 θ 求期望，而风险是对 $L(\theta,\hat\theta(X_1,\cdots,X_n))$ 同时关于 $\theta,X_1,\cdots,X_n$ 求期望. 易见

$$R_{\hat\theta} = E(R_{\hat\theta}(X_1,\cdots,X_n)). \tag{11.7}$$

损失函数的确定要根据问题的需要. 在一般的估计问题中可取 $L(\theta,a)=(\theta-a)^2$ 或者 $L(\theta,a)=|\theta-a|$. 在实际问题中，应根据实际背景来决定损失函数. 例如在本节开头所举的例子中，对日废品率 θ 若估计过高，则直接造成厂家经济损失；若估计过低，则造成信誉的损失（最终仍转化成经济损失），因此损失函数可取如下的形式：

$$L(\theta,a) = \begin{cases} h_1(a-\theta), \text{当 } a>\theta, \\ h_2(\theta-a), \text{当 } a\leqslant\theta, \end{cases}$$

其中 h_1,h_2 均为(0,1)上的非负增函数.

11.5.2 贝叶斯估计

从贝叶斯学派的观点来看，评价一个估计 $\hat\theta$ 的优劣是看它的后验风险与风险的大小. 对于一次性使用的估计量，希望其后验风险愈小愈好；对于多次重复使用的估计量，则希望其风险愈小愈好. 由此引出

定义 11.5.6 设待估参数 θ 为随机变量，$X_1,\cdots,X_n$ 为样本，$\hat\theta=\hat\theta(X_1,\cdots,X_n)$ 为 θ 的一个估计. 对给定的损失函数 $L(\theta,a)$，记 $R_{\hat\theta}$ 为 $\hat\theta$ 的风险（见 11.5.5）. 若对 θ 的任一估计 $\hat\theta^*$ 用 $R_{\hat\theta^*}$ 记其风险，有

$$R_{\hat\theta} \leqslant R_{\hat\theta^*},$$

则称 $\hat\theta$ 为**贝叶斯估计**(Bayes estimate).

贝叶斯估计为贝叶斯意义下的最优估计. 除了先验分布和样本分布之外，贝叶斯估计还与损失函数的选取有关.

定义 11.5.7 在定义 11.5.6 的记号之下. 设给定样本 $X_1=x_1,\cdots,X_n=x_n$,记 $R_{\hat{\theta}}(x_1,\cdots,x_n)$为 $\hat{\theta}$ 的后验风险(见定义 11.5.4). 对于 θ 的任一估计 $\hat{\theta}^*$ 用 $R_{\hat{\theta}^*}(x_1,\cdots,x_n)$记其后验风险有

$$R_{\hat{\theta}}(x_1,\cdots,x_n)\leqslant R_{\hat{\theta}^*}(x_1,\cdots,x_n).$$

则称 $\hat{\theta}$ 为**后验风险最小估计**.

后验风险最小估计有重要的实际和理论意义. 根据(11.2)式易见:若一个估计 $\hat{\theta}$ 对每个可能实现的样本点都是后验风险最小估计,则它必然是贝叶斯估计. 在很一般的条件下,这个结论的逆命题也成立.

定理 11.5.8 设 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$为 θ 的一个估计,其风险

$$R_{\hat{\theta}}<\infty,$$

则 $\hat{\theta}$ 为贝叶斯估计的充分必要条件是对每一可能实现的样本点,$\hat{\theta}$ 为后验风险最小估计.

定理 11.5.8 被称为是"后验风险最小原理".

下面考虑损失函数为 $L(\theta,a)=(\theta-a)^2$ 的情形. 给定样本 $X_1=x_1,\cdots,X_n=x_n$,记后验分布为 $H(\theta|x_1,\cdots,x_n)$. 这时用 a 去估计 θ 的后验风险为

$$\begin{aligned}R_a(x_1,\cdots,x_n)&=\int_{-\infty}^{\infty}(\theta-a)^2\mathrm{d}H(\theta|x_1,\cdots,x_n)\\&=a^2-2a\int_{-\infty}^{\infty}\theta\mathrm{d}H(\theta|x_1,\cdots,x_n)\\&\quad+\int_{-\infty}^{\infty}\theta^2\mathrm{d}H(\theta|x_1,\cdots,x_n)\\&=a^2-2aE[\theta|x_1,\cdots,x_n]+E[\theta^2|x_1,\cdots,x_n].\end{aligned}$$

显然,这是一个关于 a 的二项式,而为使后验风险达到最小,应有

$$a=E[\theta|x_1,\cdots,x_n].$$

定理 11.5.9 设 θ 为待估参数,$X_1,\cdots,X_n$ 为样本,则对损失函数$(\theta-a)^2$,贝叶斯估计为

$$\hat{\theta} = E[\theta | X_1, \cdots, X_n].$$

若损失函数为 $L(\theta,a)=|\theta-a|$，则用 a 去估计 θ 的后验风险为

$$R_a(x_1, \cdots, x_n) = \int_{-\infty}^{\infty} |\theta - a| \mathrm{d}H(\theta | x_1, \cdots, x_n).$$

在这种情形下，为求 R_a 的最小值点，可利用下述引理：

引理 11.5.10 设 X 为随机变量，$h(a)=E|X-a|$，则 $h(a)$ 的最小值在 X 的中位数 $m(X)$ 处达到，即

$$E|X - m(x)| = \min_a E|X - a|.$$

利用上述引理可知，对损失函数 $|\theta-a|$，θ 的后验分布的中位数 $\mu(x_1,\cdots,x_n)$ 使后验风险达到最小.

定理 11.5.11 设 θ 为待估参数，$X_1,\cdots,X_n$ 为样本，则对损失函数 $|\theta\text{-}a|$，贝叶斯估计为

$$\hat{\theta} = \mu(X_1, \cdots, X_n),$$

其中 $\mu(X_1,\cdots,X_n)$ 为 θ 关于样本 $X_1,\cdots,X_n$ 的条件中位数.

下面给出几个计算的例子.

例 11.5.12 设 θ 为一批产品中的次品率 $(0\leqslant\theta\leqslant1)$. 设 θ 的先验分布为 $[0,1]$ 区间上的均匀分布 $U[0,1]$. 现从该批产品中抽得 n 个样品，记其次品个数为 X，取损失函数 $L(\theta,a)=(\theta-a)^2$. 求 θ 的贝叶斯估计.

解 X 的分布为二项分布 $B(n,\theta)$，于是 X 与 θ 的联合频率函数为 $\binom{n}{x}\theta^x(1-\theta)^{n-x}$，$x=0,1,\cdots,n$；$0\leqslant\theta\leqslant1$. 因此，给定 $X=x$ 时，θ 的后验分布频率函数为

$$h(\theta|x) = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}\mathrm{d}\theta}$$

$$= \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}.$$

根据定理 11.5.9,贝叶斯估计为

$$\begin{aligned}\hat{\theta} &= E[\theta | X = x] \\ &= \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^1 \theta^{x+1}(1-\theta)^{n-x}\mathrm{d}\theta \\ &= \frac{x+1}{n+2}.\end{aligned}$$

这时可算得贝叶斯风险为$\frac{1}{6(n+2)}$. 而通常用 X/n 去估计 θ 的贝叶斯风险为$\frac{1}{6n}$. 一个令人感兴趣的事实是:在这个例中,当样品中无次品(或全部为次品时),不是用 0(或用 1)来估计次品率 θ,而是用$\frac{1}{n+2}\left(\text{或用}\frac{n+1}{n+2}\right)$来估计,这与通常的估计 X/n 是不同的.

例 11.5.13 设 $X_1,\cdots,X_n$ 为抽自 $N(\theta,1)$的简单样本,θ 的先验分布为 $N(0,b^2)$,损失函数为 $L(\theta,a)=(\theta-a)^2$,求 θ 的贝叶斯估计.

解 $(\theta,X_1,\cdots,X_n)$的联合密度为

$$\frac{1}{\sqrt{2\pi}b}\exp(-\theta^2/2b^2)\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left(-\sum_{i=1}^n(x_i-\theta)^2/2\right),$$

注意其指数为

$$-\frac{1}{2}\left[\theta^2/b^2+\sum_{i=1}^n(x_i-\theta)^2\right]=-\frac{1+nb^2}{2b^2}\left(\theta-\frac{nb^2\overline{x}}{1+nb^2}\right)^2+C,$$

其中 $\overline{x}=\frac{1}{n}\sum_{i=1}^n x_i$,C 为与 θ 无关的量. 易知给定样本 $X_1=x_1,\cdots,X_n=x_n$ 时,θ 的后验分布为 $N\left(\frac{nb^2\overline{x}}{1+nb^2},\frac{b^2}{1+nb^2}\right)$. 由此知 θ 的贝叶斯估计为

$$\hat{\theta}=\frac{nb^2}{1+nb^2}\overline{X},$$

其中 $\overline{X}$ 为样本均值. 贝叶斯风险为$\frac{b^2}{1+nb^2}$.

在此例中 $\overline{X}$ 是方差最小无偏估计，也是极大似然估计，而贝叶斯估计相当于将 $\overline{X}$ 乘上一个小于1的正数因子，即将 $\overline{X}$ 压缩一点. 这是因为先验分布 $N(0,b^2)$说明 θ 的取值应集中在0点附近，因此在用 $\overline{X}$ 去估计 θ 之值时应适当向0点附近收缩一下.

11.5.3 经验贝叶斯估计

贝叶斯估计的缺点是先验分布往往很难确定. 当先验分布未知或先验分布中含有未知参数时就无法找到贝叶斯估计. 罗宾斯(Robbins)于1955年提出了**经验贝叶斯方法**(empirical Bayes method)，在一定程度上克服了这个困难. 设一个总体 X 有分布 F_θ，其中 θ 为未知参数. 假设 θ 随时间流逝而波动，因此某一时刻 θ 的取值可以看成是随机的. 作为历史资料，假定我们有 m 批样本：$X_1=(X_1^{(1)},\cdots,X_{n_1}^{(1)}),\cdots,X_m=(X_1^{(m)},\cdots,X_{n_m}^{(m)})$，则这 m 批样本包含了 θ 过去取值的一些信息. 设当前的样本为 $X_{m+1}=(X_1^{(m+1)},\cdots,X_{n_{m+1}}^{(m+1)})$，现在要用这些历史样本与当前样本来估计 θ 当前的值.

方法 11.5.14 经验贝叶斯方法 该方法的基本思想是：利用历史样本 $X_1,\cdots,X_m$ 去估计 θ 的先验分布，设得到的先验分布的估计为 $\hat{H}(\theta)$. 把 $\hat{H}(\theta)$当作 θ 的真正先验分布，与当前样本 X_{m+1}合在一起可以得到一个贝叶斯估计.

显然，先验分布的估计 $\hat{H}(\theta)$与真正的先验分布 $H(\theta)$不必完全相同，因此用经验贝叶斯方法得到的估计不一定是贝叶斯估计. 此外，$\hat{H}(\theta)$可以有不同的形式(用不同的方法去估计先验分布 $H(\theta)$)，因此由同一批历史样本和当前样本可以得到不同的经验贝叶斯估计.

下面介绍两种从实用的观点上看可行的经验贝叶斯估计方法.

方法 11.5.15 第一种经验贝叶斯估计方法 假定根据过去

的经验已知 θ 的先验分布有给定的分布形式 $H_\lambda(\theta)$，其中 λ 为未知参数. 因此，可以用历史样本 $X_1,\cdots,X_m$，对 λ 产生一个估计 $\hat{\lambda}=\hat{\lambda}(X_1,\cdots,X_m)$，而先验分布的估计为 $H_{\hat{\lambda}}(\theta)$. 再利用 $H_{\hat{\lambda}}(\theta)$ 作为先验分布，与当前样本 X_{m+1} 一起来构成贝叶斯估计.

例 11.5.16 设总体 X 服从 $N(\theta,1)$ 分布，损失函数 $L(\theta,a)=(\theta-a)^2$，θ 的先验分布为 $N(0,\sigma^2)$，其中 σ^2 为未知参数. 由此可以算得 X 的无条件分布为 $N(0,1+\sigma^2)$（请读者自行验证）. 假定历史样本为 $X_1,\cdots,X_n$，当前样本为 X_{n+1}. 于是由历史样本可得 σ^2 的一个估计为

$$\hat{\sigma}^2=\sum_{i=1}^{n}X_i^2-1,$$

于是 θ 的先验分布的估计为 $N(0,\hat{\sigma}^2)$. 根据例 11.5.13 知贝叶斯估计为

$$\hat{\theta}=\frac{\hat{\sigma}^2}{1+\hat{\sigma}^2}X_{n+1}.$$

方法 11.5.17 第二种经验贝叶斯方法 假定先验分布的形式事先也无法确定，则问题变得更为困难. 根据 θ 在总体分布中的特征性质，我们可以用通常的估计方法，从历史样本 $X_1=\left(X_1^{(1)},\cdots,X_{n_1}^{(1)}\right),\cdots,X_m=\left(X_1^{(m)},\cdots,X_{n_m}^{(m)}\right)$ 出发对 θ 得到 m 个估计 $\hat{\theta}_1=\hat{\theta}(X_1),\cdots,\hat{\theta}_m=\hat{\theta}(X_m)$. 将这 m 个估计看成是 θ 的一个容量为 m 的简单样本，便可得一个经验分布 H_m（见定义 14.2.5）. 用 H_m 作为先验分布的估计与当前样本 X_{m+1} 合在一起，便可得到一个贝叶斯估计.

例 11.5.18 设总体分布有频率函数 $f(x|\theta)$，其中 θ 为未知参数. 设 $\hat{\theta}_1,\cdots,\hat{\theta}_m$ 为用 m 个历史样本得到的关于 θ 的 m 个估计，当前样本为 $X_1,\cdots,X_n$，损失函数取 $(\theta-a)^2$. 于是 θ 与当前样本有联合的（经验）频率函数：

$$\frac{1}{m}\sum_{i=1}^{m}I_{(\theta=\hat{\theta}_i)}\prod_{j=1}^{n}f(x_j|\theta).$$

于是 θ 的后验频率函数为

$$\hat{h}(\theta)=\frac{\sum_{i=1}^{m} I_{(\theta=\hat{\theta}_i)} \prod_{j=1}^{n} f(x_j|\theta)}{\sum_{i=1}^{m} \prod_{j=1}^{n} f(x_j|\hat{\theta}_i)},$$

其中 $I_{(\theta=\hat{\theta}_i)}$ 为示性函数，即

$$I_{(\theta=\hat{\theta}_i)}=\begin{cases}1, & \text{当 } \theta=\hat{\theta}_i, \\ 0, & \text{当 } \theta\neq\hat{\theta}_i.\end{cases}$$

此时贝叶斯估计为

$$\hat{\theta}=\int\theta\hat{h}(\theta)\mathrm{d}\theta=\frac{\sum_{i=1}^{m}\hat{\theta}_i \prod_{j=1}^{n} f(x_j|\hat{\theta}_i)}{\sum_{i=1}^{m} \prod_{j=1}^{n} f(x_j|\hat{\theta}_i)}.$$

统计理论关心的一个问题是：经验贝叶斯估计与理论的贝叶斯估计的关系. 这方面有下述概念.

定义 11.5.19 设 $X_1,\cdots,X_m$ 为历史样本，X_{m+1} 为当前样本，经验贝叶斯估计为 $\hat{\theta}_{m+1}=\hat{\theta}(X_{m+1}|X_1,\cdots,X_m)$. 用 R 记理论上的贝叶斯风险值，R_{m+1} 为 $\hat{\theta}_{m+1}$ 的（无条件）风险. 若

$$\lim_{m\to\infty} R_{m+1}=R.$$

则称 $\hat{\theta}_{m+1}$ 为**渐近最优经验贝叶斯估计**.

渐近最优贝叶斯估计的思想是：当历史样本足够多时，可以任意精确地再现 θ 的先验分布，因而可以任意精确地逼近贝叶斯风险.

12 区间估计

12.1 置信区间

12.1.1 基本概念

在点估计问题中，对一个未知参数 θ 用一个样本的函数 $\hat{\theta}$ 去估计它. 当给定样本值时，$\hat{\theta}$ 是一个点. 估计的精度用 $\hat{\theta}$ 关于 θ 的均方误差(对无偏估计来说是方差)来衡量. 而区间估计是从另一个角度来解决估计 θ 的问题.

定义 12.1.1 设 θ 为一维未知参数，$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 为两个统计量，满足 $\hat{\theta}_1 \leqslant \hat{\theta}_2$. 用区间$[\hat{\theta}_1, \hat{\theta}_2]$去估计 θ 可能存在的范围，称为 θ 的一个**区间估计**(interval estimation).

同构造点估计一样，构造区间估计也要遵循一定的准则. 由不同的准则出发便得到不同的区间估计. 目前在应用中最广泛的构造区间估计的理论和方法是源于内曼(Neymann)的置信区间的概念.

定义 12.1.2 设$[\hat{\theta}_1, \hat{\theta}_2]$为 θ 的一个区间估计，若对给定的正数 $1-\alpha$ 及 θ 的任一可能值 θ' 有

$$P_{\theta'}(\hat{\theta}_1 \leqslant \theta' \leqslant \hat{\theta}_2) \geqslant 1-\alpha,$$

则称$[\hat{\theta}_1, \hat{\theta}_2]$为 θ 的一个**置信水平**(confidence level)为 $1-\alpha$ 的**置信区间**(confidence inteval).

注

(1) 在定义 12.1.2 中，$P_{\theta'}$是指由 $\theta=\theta'$ 时的样本分布计算的概率. 在某些特定的场合下，P_θ 可能与 θ 无关.

(2) 在置信区间的概念中，θ 虽然未知，但被看成是固定常数，

而区间端点 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 作为样本的函数是随机变量. 因此事件$\{\theta\in[\hat{\theta}_1,\hat{\theta}_2]\}$的确切含义是"随机区间$[\hat{\theta}_1,\hat{\theta}_2]$包含常数 θ",但习惯上也常说"θ 落在区间$[\hat{\theta}_1,\hat{\theta}_2]$内".

置信区间的概念有一个频率解释:对 θ 取任何可能值,在固定的置信水平 $1-\alpha$ 下多次使用同一个区间估计(样本值虽然每次不同,但构造区间估计的方法不变),则 θ 落在$[\hat{\theta}_1,\hat{\theta}_2]$内的频率大致为 $1-\alpha$,而 θ 落在$[\hat{\theta}_1,\hat{\theta}_2]$之外的频率大致为 α.

置信水平 $1-\alpha$ 的大小反映了置信区间的可靠程度,在实用中常取$\alpha=0.90,0.95$ 或 0.99 等值. 置信区间的平均长度 $2d=E(\hat{\theta}_2-\hat{\theta}_1)$的大小则反映了置信区间的精度. 通常,在固定样本容量 n 时,置信水平越高($1-\alpha$ 越大),则置信区间的精度就越低(区间的平均长度越大);反之亦然. 因此,在给定的置信水平下,为取得足够的精度就必须使样本的容量足够大.

例 12.1.3 设总体分布为 $N(\mu,\sigma_0^2)$,其中 μ 为未知参数,σ_0^2 已知,$X_1,\cdots,X_n$ 为简单样本. 求 μ 的置信水平为 0.95 的置信区间.

由于样本均值 $\overline{X}$ 为 μ 的一个理想的点估计,因此考虑以 $\overline{X}$ 为中心来构造 μ 的区间估计. 由 $\overline{X}\sim N\left(\mu,\frac{\sigma_0^2}{n}\right)$ 知

$$U=\frac{\sqrt{n}\,(\overline{X}-\mu)}{\sigma_0}\sim N(0,1),$$

因此(注意 U 的分布与 μ 无关!)

$$P(-1.96\leqslant U\leqslant 1.96)=0.95.$$

而事件$\{-1.96\leqslant U\leqslant 1.96\}=\left\{-1.96\leqslant\frac{\sqrt{n}\,(\overline{X}-\mu)}{\sigma_0}\leqslant 1.96\right\}$可通过不等式变换转化为一个等价的事件

$$\left\{\overline{X}-1.96\frac{\sigma_0}{\sqrt{n}}\leqslant\mu\leqslant\overline{X}+1.96\frac{\sigma_0}{\sqrt{n}}\right\}.$$

因此取 $\hat{\mu}_1=\overline{X}-1.96\dfrac{\sigma_0}{\sqrt{n}}$，$\hat{\mu}_2=\overline{X}+1.96\dfrac{\sigma_0}{\sqrt{n}}$，则有

$$P\{\mu\in[\hat{\mu}_1,\hat{\mu}_2]\}=0.95.$$

因此$[\hat{\mu}_1,\hat{\mu}_2]$为 μ 的一个置信水平 $1-\alpha$ 的置信区间. 区间长度为常量：$2d=3.92\dfrac{\sigma_0}{\sqrt{n}}$. 假若事先给定 d，则应取样本容量 $n\geqslant(1.96\sigma_0/d)^2$.

在此例中，置信水平 $1-\alpha$ 的置信区间不是唯一的. 例如，通过查正态分布函数表，容易得到 $P(U<-2.33)=0.01$ 而 $P(U>1.76)=0.04$. 因此 $P(-2.33\leqslant U\leqslant 1.76)=0.95$. 相应得到的置信区间为$\left[\overline{X}-1.76\dfrac{\sigma_0}{\sqrt{n}},\overline{X}+2.33\dfrac{\sigma_0}{\sqrt{n}}\right]$. 这时置信区间不是以 $\overline{X}$ 为中心且区间长度为 $4.09\dfrac{\sigma_0}{\sqrt{n}}$，大于以 $\overline{X}$ 为中心的区间长度. 容易证明，在所有由 U 所导出的 μ 的置信区间中，以 $\overline{X}$ 为中心的区间长度最小，因此最为常用. 但在某些特定的场合，例如当 μ 估计得过小会产生某种不良后果时，则后一类型的置信区间也可能用到. 总之，有时根据问题的需要，只需估计未知参数的下限或上限.

定义 12.1.4 设 θ 为未知参数，$1-\alpha$ 为给定的置信水平，若统计量 $\underline{\theta}$ 和 $\overline{\theta}$ 分别满足

$$P_\theta(\underline{\theta}\leqslant\theta)\geqslant 1-\alpha,\forall\ \theta,$$

$$P_\theta(\theta\leqslant\overline{\theta})\geqslant 1-\alpha,\forall\ \theta,$$

则称 $\underline{\theta}(\overline{\theta})$为 θ 的置信水平 $1-\alpha$ 的**置信下(上)限**(confidence lower (upper) limit).

推论 12.1.5 设 $\alpha_0+\alpha_1=\alpha$，$\underline{\theta}$ 为 θ 的置信水平 $1-\alpha_0$ 的置信下限，$\overline{\theta}$ 为 θ 的置信水平 $1-\alpha_1$ 的置信上限，则$[\underline{\theta},\overline{\theta}]$为 θ 的置信水平 $1-\alpha$ 的置信区间.

由例 12.1.3 还可看出，构造 μ 的区间估计涉及到统计量 $U=$

$\sqrt{n}(\overline{X}-\mu)/\sigma_0$，它有两条重要性质．第一，$U$ 的分布与未知参数 μ 无关；第二，U 是 μ 的严格单调函数，因此，形如$\{z_1\leqslant U\leqslant z_2\}$的事件可等价地转化为形如$\{\hat{\mu}_1\leqslant\mu\leqslant\hat{\mu}_2\}$．因此，寻找有这两条性质的统计量就成为构造区间估计的一个重要依据，在多数场合下是行之有效的(但不是在任何场合下都有效)．

12.1.2 单正态总体参数的置信区间

设总体分布为 $N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 为简单样本，$1-\alpha$ 为给定的置信水平．

方法 12.1.6 σ^2 已知时，μ 的置信区间．

这个情况已在例 12.1.3 中讨论过，记 $U=\sqrt{n}(\overline{X}-\mu)/\sigma$，则 $U\sim N(0,1)$．取 $\alpha_0>0$，$\alpha_1>0$，$\alpha_0+\alpha_1=\alpha$，则 $P(U>z_{\alpha_1})=\alpha_1$，$P(U<-z_{\alpha_0})=\alpha_0$，其中 z_{α_0}和 z_{α_1}分别为 $N(0,1)$的上 α_0 和上 α_1 分位数，可在正态分布函数表(附表 1)中查到．因此

$$P(-z_{\alpha_0}\leqslant U\leqslant z_{\alpha_1})=1-(\alpha_0+\alpha_1)=1-\alpha.$$

将关于 U 的不等式变到关于 μ 的不等式(见例 12.1.3)，有

$$P\left(\overline{X}-z_{\alpha_1}\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\overline{X}+z_{\alpha_0}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha.$$

因此$\left[\overline{X}-z_{\alpha_1}\dfrac{\sigma}{\sqrt{n}},\overline{X}+z_{\alpha_0}\dfrac{\sigma}{\sqrt{n}}\right]$为 μ 的一个置信水平 $1-\alpha$ 的置信区间．当 $\alpha_0=\alpha_1=\alpha/2$ 时，区间长度为最短，即 $2z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$．当取 $\alpha_0=0(\alpha_1=\alpha)$时，便得到 μ 的置信水平 $1-\alpha$ 的置信下限 $\underline{\mu}=\overline{X}-z_\alpha\dfrac{\sigma}{\sqrt{n}}$，同理可得 μ 的相应水平的置信上限为 $\overline{\mu}=\overline{X}+z_\alpha\dfrac{\sigma}{\sqrt{n}}$．

方法 12.1.7 σ^2 未知时，μ 的置信区间

此时统计量 $U=\sqrt{n}(\overline{X}-\mu)/\sigma$ 仍服从 $N(0,1)$分布，但 U 不只是 μ 的函数，还是未知参数 σ^2 的函数．因此不能用 U 来构造 μ

的置信区间. 英国统计学家哥塞特(Gosset)(笔名 Student)提出用

$$T = \sqrt{n}(\overline{X} - \mu)/S.$$

来构造 μ 的区间估计,其中 $S=\sqrt{\frac{1}{n-1}\sum_{1}^{n}(X_i-\overline{X})^2}$,称为样本标准差. T 就是在 U 中用 S 代替 σ 所得到的. 由定理 6.4.21 知,$T\sim t(n-1)$:自由度为 $n-1$ 的 t 分布. 因此对 $\alpha_0+\alpha_1=\alpha$,有

$$P(-t_{\alpha_0}(n-1) \leqslant T \leqslant t_{\alpha_1}(n-1)) = 1-\alpha,$$

其中 $t_{\alpha_0}(n-1)$和 $t_{\alpha_1}(n-1)$分别为 $t(n-1)$的上 α_0 和上 α_1 分位数. 将关于 T 的不等式变换到关于 μ 的不等式,有

$$P\left(\overline{X} - t_{\alpha_1}(n-1)\frac{S}{\sqrt{n}} \leqslant \mu \leqslant \overline{X} + t_{\alpha_0}(n-1)\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

因此,μ 的置信水平 $1-\alpha$ 的一个置信区间为 $\left[\overline{X}-t_{\alpha_1}(n-1)\frac{S}{\sqrt{n}},\ \overline{X}+t_{\alpha_0}(n-1)\frac{S}{\sqrt{n}}\right]$. 当 $\alpha_0=\alpha_1=\alpha/2$ 时区间长度最短,为 $2d=2t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$,$d$ 是一个随机变量. 相应置信水平的置信上限为 $\overline{X}+t_{\alpha}(n-1)\frac{S}{\sqrt{n}}$,置信下限为 $\overline{X}-t_{\alpha}(n-1)\frac{S}{\sqrt{n}}$.

方法 12.1.8 μ 已知时,σ^2 的置信区间

仿照由 $\overline{X}$ 出发导出均值 μ 的置信区间的方法,此处用 σ^2 的极大似然估计 $\hat{\sigma}^2=\frac{1}{n}\sum_{1}^{n}(X_i-\mu)^2$ 来导出 σ^2 的置信区间. 由 χ^2 分布的定义容易推出

$$K^2 = n\hat{\sigma}^2/\sigma^2 = \sum_{1}^{n}(X_i-\mu)^2/\sigma^2 \sim \chi^2(n).$$

因此,对 $\alpha_0+\alpha_1=\alpha$ 有

$$P\left(\chi^2_{1-\alpha_0}(n) \leqslant K^2 \leqslant \chi^2_{\alpha_1}(n)\right) = 1-\alpha,$$

其中 $\chi^2_{1-\alpha_0}(n)$和 $\chi^2_{\alpha_1}(n)$分别为 $\chi^2(n)$的上 $1-\alpha_0$ 和 α_1 分位数,可通

过查 χ^2 分布表(附表 2)来得到. 将事件$\{\chi^2_{1-\alpha_0}(n)\leqslant K^2\leqslant\chi^2_{\alpha_1}(n)\}$作等价变形得$\left\{\frac{n\hat{\sigma}^2}{\chi^2_{\alpha_1}(n)}\leqslant\sigma^2\leqslant\frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha_0}(n)}\right\}$,这就给出了 σ^2 的一个置信水平 $1-\alpha$ 的置信区间. 在实用中常令 $\alpha_0=\alpha_1=\alpha/2$,但不同于估计 μ 的场合,这样做除了方便之外并没有什么特别的意义. 相应置信水平的置信上限为$\overline{\sigma^2}=\frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha}(n)}$,置信下限为$\underline{\sigma^2}=\frac{n\hat{\sigma}^2}{\chi^2_{\alpha}(n)}$.

方法 12.1.9 μ 未知时,σ^2 的置信区间

此时,σ^2 的极大似然估计为 $S^2=\frac{1}{n-1}\sum_1^n(X_i-\overline{X})^2$,且 $K^2=(n-1)S^2/\sigma^2\sim\chi^2(n-1)$(见定理 6.4.17). 因此,$\sigma^2$ 的水平 $1-\alpha$的置信区间有形式

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha_1}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha_0}(n-1)}\right].$$

其中 $\alpha_0+\alpha_1=\alpha$. 相应水平的置信下限为$(n-1)S^2/\chi^2_\alpha(n-1)$,置信上限为$(n-1)S^2/\chi^2_{1-\alpha}(n-1)$.

12.1.3 双正态总体均值差与方差比的置信区间

设 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$ 为分别取自两个总体 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$的简单样本. 记 $\delta=\mu_1-\mu_2$, $\lambda=\sigma_1^2/\sigma_2^2$. 设 $1-\alpha$ 为给定的置信水平.

方法 12.1.10 σ_1^2 和 σ_2^2 已知时,δ 的置信区间

记 $\overline{X}=\frac{1}{n}\sum_1^n X_i$,$\overline{Y}=\frac{1}{m}\sum_1^m Y_j$,则 $\overline{X}-\overline{Y}$ 为 δ 的方差最小无偏估计,且

$$\overline{X}-\overline{Y}\sim N\left(\delta,\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right),$$

故

$$U=\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}\sim N(0,1).$$

取 $\alpha_0+\alpha_1=\alpha$,于是

$$P\{-z_{\alpha_0}\leqslant U\leqslant z_{\alpha_1}\}=1-\alpha.$$

其中 z_{α_0} 和 z_{α_1} 分别为 $N(0,1)$的上 α_0 和上 α_1 分位数. 事件 $\{-z_{\alpha_0}\leqslant U\leqslant z_{\alpha_1}\}$的另一等价形式为$\left\{\overline{X}-\overline{Y}-z_{\alpha_1}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\leqslant\delta\leqslant\overline{X}-\overline{Y}+z_{\alpha_0}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right\}$. 这就给出了 $\delta=\mu_1-\mu_2$ 的一个置信水平 $1-\alpha$ 的置信区间. 当 $\alpha_1=\alpha$ 或 $\alpha_0=\alpha$ 时,置信区间的左右端点分别给出了 δ 的置信水平 $1-\alpha$ 的置信下限和置信上限.

方法 12.1.11 σ_1^2 与 σ_2^2 未知,但已知 $\sigma_1^2=\sigma_2^2$ 时,δ 的置信区间

此时记 $\sigma^2=\sigma_1^2=\sigma_2^2$,则 $U=\sqrt{\frac{nm}{n+m}}(\overline{X}-\overline{Y}-\delta)/\sigma$ 仍然服从 $N(0,1)$分布,但 σ 未知. 此时可用

$$S_p=\left(\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}\right)^{\frac{1}{2}}$$

去代替 σ,得到一个新的统计量

$$T=\sqrt{\frac{nm}{n+m}}(\overline{X}-\overline{Y}-\delta)/S_p.$$

其中 $S_X^2=\frac{1}{n-1}\sum_1^n(X_i-\overline{X})^2$,$S_Y^2=\frac{1}{m-1}\sum_1^m(Y_j-\overline{Y})^2$,分别是用 X 样本和 Y 样本构造的 σ^2 的无偏估计. 因此 S_p^2 是 σ^2 的无偏估计. 由定理 6.4.23 知 $T\sim t(n+m-2)$. 因此对 $\alpha_0+\alpha_1=\alpha$.

$$P\{-t_{\alpha_0}(n+m-2)\leqslant T\leqslant t_{\alpha_1}(n+m-2)\}=1-\alpha.$$

由此给出 δ 的一个置信水平 1-α 的置信区间为

$$\left[\overline{X}-\overline{Y}-t_{\alpha_1}(n+m-2)S_p\Big/\sqrt{\frac{nm}{n+m}},\right.$$
$$\left.\overline{X}-\overline{Y}+t_{\alpha_0}(n+m-2)S_p\Big/\sqrt{\frac{nm}{n+m}}\right],$$

在实际问题中常取 $\alpha_0=\alpha_1=\alpha/2$. 当 $\alpha_1=\alpha$ 或 $\alpha_0=\alpha$ 时，区间的左右端点分别给出了置信水平为 $1-\alpha$ 的置信下限和置信上限.

方法 12.1.12 σ_1^2 与 σ_2^2 未知且不等时，δ 的置信区间

在前两种情形中(即 σ_1^2 与 σ_2^2 已知和 $\sigma_1^2=\sigma_2^2$ 但未知)，构造 $\delta=\mu_1-\mu_2$ 的区间估计所依据的统计量 U 和 T 有一个共同的特点，即它们关于样本 X 和样本 Y 分别是对称的，即不论对 $X_1,\cdots,X_n$ 或 $Y_1,\cdots,Y_m$ 作什么样的重排，U 和 T 的值是相同的. 但当 σ_1^2 与 σ_2^2 未知且不必相等时，要构造这样的统计量且使其分布与未知参数无关是不可能的. 下面介绍两种构造区间估计的方法.

首先，不失一般性地假定 $n\leqslant m$，有

定理 12.1.13 设 $W_i=X_i-\sum\limits_{j=1}^{m}c_{ij}Y_j, i=1,\cdots,n$，其中

$$c_{ij}=\begin{cases}\dfrac{\sqrt{n}}{m}-\dfrac{1}{\sqrt{nm}}+\dfrac{1}{m}, & \text{当 } j=i,\\[2ex] -\dfrac{1}{\sqrt{nm}}+\dfrac{1}{m}, & \text{当 } j\neq i, j\leqslant n,\\[2ex] \dfrac{1}{m}. & \text{当 } j>n.\end{cases}$$

则 $W_1,\cdots,W_n$ 独立同分布 $N(\delta,\sigma^2)$，其中 $\delta=\mu_1-\mu_2$，$\sigma^2=\sigma_1^2+\dfrac{n}{m}\sigma_2^2$.

利用此定理，令

$$T'=\frac{\sqrt{n}\,(\overline{W}-\delta)}{S_W},\tag{12.1}$$

其中 $\overline{W}=\frac{1}{n}\sum_{1}^{n}W_i, S_W^2=\frac{1}{n-1}\sum_{1}^{n}(W_i-\overline{W})^2$，则 $T'\sim t(n-1)$（见定理 6.4.21）. 因此类似单正态总体中方差未知时，均值的置信区间的构造法得到 δ 的一个置信水平 $1-\alpha$ 的置信区间为

$$\left[\overline{W}-t_{\alpha_1}(n-1)\frac{S_W}{\sqrt{n}},\overline{W}+t_{\alpha_0}(n-1)\frac{S_W}{\sqrt{n}}\right],$$

其中 $\alpha_1+\alpha_0=\alpha$.

这样构造的区间估计的缺点是它关于样本 Y 不对称. 但理论上可以证明，这个区间估计的精度还是不错的. 在实际使用时应注意：由于样本的排列往往有某种"自然"的顺序，例如在寿命试验中，样本的记录往往是依寿命从小到大的顺序排列的，因此在构造 $W_i, 1\leqslant i\leqslant n$ 之前应对 Y 样本（即容量大的样本）做一随机的重排.

另一个思路是：保持统计量关于样本的对称性，而分布是近似的. 有

定理 12.1.14 设

$$T''=(\overline{X}-\overline{Y}-\delta)\Bigg/\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}. \tag{12.2}$$

其中 $S_X^2=\frac{1}{n-1}\sum_{1}^{n}(X_i-\overline{X})^2, S_Y^2=\frac{1}{m-1}\sum_{1}^{m}(Y_j-\overline{Y})^2$.

则 T'' 的分布近似地与自由度 ν 的 t 分布相同，其中

$$\nu=\left(\frac{\lambda}{n}+\frac{1}{m}\right)^2\Bigg/\left(\frac{\lambda^2}{n^2(n-1)}+\frac{1}{m^2(m-1)}\right), \lambda=\sigma_1^2/\sigma_2^2.$$

利用此定理，令 $\hat{\lambda}=S_X^2/S_Y^2$，将 $\hat{\lambda}$ 代入 ν 的表达式中得到 $\hat{\nu}$，可以近似地认为

$$T''\sim t(\hat{\nu}).$$

由此得到 δ 的一个近似置信水平为 $1-\alpha$ 的置信区间为

$$\left[\overline{X}-\overline{Y}-t_{\alpha_0}(\hat{\nu})\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}},\overline{X}-\overline{Y}+t_{\alpha_0}(\hat{\nu})\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}\right].$$

这两种区间估计的实际效果往往相差不大.

方法 12.1.15 μ_1 和 μ_2 均已知时，$\lambda=\sigma_1^2/\sigma_2^2$ 的置信区间

此时，$\hat{\sigma}_1^2=\frac{1}{n}\sum_1^n(X_i-\mu_1)^2$ 和 $\hat{\sigma}_2^2=\frac{1}{m}\sum_1^n(Y_j-\mu_2)^2$ 分别为 σ_1^2 和 σ_2^2 的最小方差无偏估计，因此考虑由统计量

$$F=\frac{\hat{\sigma}_1^2/\sigma_1^2}{\hat{\sigma}_2^2/\sigma_2^2}=\lambda^{-1}\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}.$$

出发来构造 λ 的置信区间. 由定义 6.4.24 易知 $F\sim F(n,m)$：自由度为 (n,m) 的 F 分布. 因此

$$P\left(F_{1-\alpha_0}(n,m)\leqslant F\leqslant F_{\alpha_1}(n,m)\right)=1-\alpha,$$

其中 $\alpha_0+\alpha_1=\alpha$，$F_{1-\alpha_0}(n,m)$ 和 $F_{\alpha_1}(n,m)$ 分别为 $F(n,m)$ 的上 $1-\alpha_0$ 和上 α_1 分位数，可由 F 分布表(附表 4)上查得. 因此，λ 的置信水平 $1-\alpha$ 的置信区间为

$$\left[\frac{\hat{\sigma}_1^2/\hat{\sigma}_2^2}{F_{\alpha_1}(n,m)},\frac{\hat{\sigma}_1^2/\hat{\sigma}_2^2}{F_{1-\alpha_0}(n,m)}\right].$$

当 $\alpha_1=\alpha$ 或 $\alpha_0=\alpha$ 时，区间的左端点或右端点给出置信下限或置信上限.

方法 12.1.16 μ_1 和 μ_2 均未知时，$\lambda=\sigma_1^2/\sigma_2^2$ 的置信区间

此时，令

$$F=\lambda^{-1}\frac{S_X^2}{S_Y^2},$$

则 $F\sim F(n-1,m-1)$，λ 的水平 $1-\alpha$ 的置信区间为

$$\left[\frac{S_X^2/S_Y^2}{F_{\alpha_1}(n-1,m-1)},\frac{S_X^2/S_Y^2}{F_{1-\alpha_0}(n-1,m-1)}\right],$$

其中 $\alpha_0+\alpha_1=\alpha$. 类似前面的情形，可得到置信上、下限.

12.1.4 渐近置信区间

在(单或双)正态总体的区间估计中，我们发现一个令人满意

的事实，即可以构造一个待估参数和样本的函数 h，满足以下两个条件：第一，h 的分布与未知参数无关，且这分布（如正态，χ^2，t 和 F 分布）的概率值有表可查；第二，h 是待估参数的单调函数，因此形如 $\{h_1 \leqslant h \leqslant h_2\}$ 的区间可以转化为待估参数的区间.

但对许多非正态总体的区间估计问题，上述条件不再存在. 需要采用别的办法. 当样本容量较大时，一个简便有效的办法是利用中心极限定理作渐近置信区间估计.

例 12.1.17 设总体 X 服从 0-1 分布，$P(X=1)=p$，$P(X=0)=1-p$，$X_1,\cdots,X_n$ 为简单样本，要构造 p 的置信区间. 这样的问题在产品抽样检验中经常碰到. 显然从 $\overline{X}=\frac{1}{n}\sum_1^n X_i$ 出发来构造 p 的区间估计是自然的. 根据中心极限定理（定理 10.1.1 或定理 10.1.9）有

$$\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}} \xrightarrow{\text{d}} N(0,1)\text{，当 } n \to \infty.$$

因此，当 n 比较大时，有

$$P\left\{-z_{\frac{\alpha}{2}} \leqslant \frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}} \leqslant z_{\frac{\alpha}{2}}\right\} \approx 1-\alpha,$$

其中 $z_{\frac{\alpha}{2}}$ 为 $N(0,1)$ 的上 $\alpha/2$ 分位数. 不等式

$$\left|\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}}\right| \leqslant z_{\frac{\alpha}{2}}$$

容易变形为二项不等式 $ap^2+bp+c\leqslant 0$，其中 $a=n+(z_{\frac{\alpha}{2}})^2$，$b=-\left(z_{\frac{\alpha}{2}}^2+2n\overline{X}\right)$，$c=n\overline{X}^2$. 由于 $a>0$ 且判别式 $\Delta=b^2-4ac>0$，可解得 ap^2+bp+c 的两个根 $\hat{p}_1<\hat{p}_2$，因此上述不等式的等价形式为 $\hat{p}_1\leqslant p\leqslant\hat{p}_2$，这就给出了 p 的一个渐近置信水平 $1-\alpha$ 的置信区间 $[\hat{p}_1,\hat{p}_2]$.

例 12.1.18 设总体分布为泊松分布 $P(\lambda)$，其中 λ 为未知参

数，$X_1,\cdots,X_n$ 为简单样本. 对给定的水平 $1-\alpha$，由中心极限定理 10.1.9 有

$$\frac{\sqrt{n}\,(\overline{X}-\lambda)}{\sqrt{\lambda}} \xrightarrow{\mathrm{d}} N(0,1)\text{，当 } n\to\infty.$$

因此，当 n 较大时，有

$$P\left\{\left|\frac{\sqrt{n}\,(\overline{X}-\lambda)}{\sqrt{\lambda}}\right|\leqslant z_{\frac{\alpha}{2}}\right\}\approx 1-\alpha,$$

其中 $z_{\frac{\alpha}{2}}$ 为 $N(0,1)$ 的上 $\alpha/2$ 分位数. 而不等式 $\left|\frac{\sqrt{n}\,(\overline{X}-\lambda)}{\sqrt{n}}\right|\leqslant z_{\frac{\alpha}{2}}$ 可表达为二次不等式 $a\lambda^2+b\lambda+c\leqslant 0$，其中 $a=n,b=-(2n\,\overline{X}+z_{\frac{\alpha}{2}}^2),c=n\,\overline{X}^2$，且 $\Delta=b^2-4ac>0$. 由此可解得 $a\lambda^2+b\lambda+c=0$ 的两个根 $\hat{\lambda}_1<\hat{\lambda}_2$. 这样就给出 λ 的一个渐近置信水平为 $1-\alpha$ 的置信区间 $[\hat{\lambda}_1,\hat{\lambda}_2]$.

一般，对简单样本 $X_1,\cdots,X_n$，用 $L(\theta)$ 记它的似然函数，则在很一般的条件下有

$$\frac{\partial\log L(\theta)}{\partial\theta}\Bigg/\left[E\left(\frac{\partial\log L(\theta)}{\partial\theta}\right)^2\right]^{\frac{1}{2}}\xrightarrow{\mathrm{d}} N(0,1). \qquad (12.3)$$

因此，假定不等式 $\left|\frac{\partial\log L}{\partial\theta}\Big/\left[E\left(\frac{\partial\log L}{\partial\theta}\right)^2\right]^{\frac{1}{2}}\right|\leqslant z_{\frac{\alpha}{2}}$ 可以表成 θ 的一个不等式 $\hat{\theta}_1\leqslant\theta\leqslant\hat{\theta}_2$，则 $[\hat{\theta}_1,\hat{\theta}_2]$ 就给出了一个渐近水平 $1-\alpha$ 的置信区间. 如果不行，则在很多场合下有 $\frac{\partial\log L}{\partial\theta}\Big/\left[E\left(\frac{\partial\log L}{\partial\theta}\right)^2\right]^{\frac{1}{2}}=g(X_1,\cdots,X_n,\theta)/h(\theta)$，其中 g 为 θ 的单调函数，而 $h(\theta)$ 存在一个相合估计 $\hat{h}$，则有

$$g(X_1,\cdots,X_n,\theta)/\hat{h}\xrightarrow{\mathrm{d}} N(0,1). \qquad (12.4)$$

此时由 g 关于 θ 的单调性，很容易将不等式 $|g/\hat{h}|\leqslant z_{\frac{\alpha}{2}}$ 转化为一个关于 θ 的不等式 $\hat{\theta}_1\leqslant\theta\leqslant\hat{\theta}_2$. 这样也给出一个渐近水平 $1-\alpha$ 的置

信区间$[\hat{\theta}_1, \hat{\theta}_2]$. 例如，在例 12.1.16 中用 $\sqrt{n}(\overline{X}-p)/\sqrt{\overline{X}(1-\overline{X})}$ 代替 $\sqrt{n}(\overline{X}-p)/\sqrt{p(1-p)}$，就可以得到 p 的一个渐近水平 $1-\alpha$ 的置信区间：

$$\left[\overline{X}-z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \overline{X}+z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right].$$

12.2 贝叶斯区间估计

在置信区间的理论中，未知参数 θ 是被看作为常数. 而根据贝叶斯学派的观点，未知参数 θ 也是随机变量（见 11.5）. 这时区间估计的形式和含义与置信区间的概念有所不同.

设当给定样本 $X_1=x_1,\cdots,X_n=x_n$ 时，θ 有后验分布函数 $F(\theta|x_1,\cdots,x_n)$. 对给定的置信水平 $1-\alpha$，取 $\alpha_0+\alpha_1=\alpha$，设 $\hat{\theta}_1=g_1(x_1,\cdots,x_n)$，$\hat{\theta}_2=g_2(x_1,\cdots,x_n)$，满足

$$P(\theta<\hat{\theta}_1|X_1=x_1,\cdots,X_n=x_n)=F(\hat{\theta}_1|x_1,\cdots,x_n)\leqslant\alpha_0,$$
$$P(\theta>\hat{\theta}_2|X_1=x_1,\cdots,X_n=x_n)=1-F(\hat{\theta}_2|x_1,\cdots,x_n)\leqslant\alpha_1,$$

则显然有

$$\begin{aligned}&P(\hat{\theta}_1\leqslant\theta\leqslant\hat{\theta}_2|X_1=x_1,\cdots,X_n=x_n)\\&=F(\hat{\theta}_2|x_1,\cdots,x_n)-F(\hat{\theta}_1|x_1,\cdots,x_n)\geqslant1-\alpha.\end{aligned}\quad(12.5)$$

由于上式对任何给定的样本值都成立. 因此，把样本看成随机变量，对 θ 和样本的联合分布求概率就有

$$P(\hat{\theta}_1\leqslant\theta\leqslant\hat{\theta}_2)\geqslant1-\alpha.\quad(12.6)$$

这样就给出了贝叶斯意义下置信水平为 $1-\alpha$ 的区间估计.

贝叶斯区间估计的形式和含义与上一节介绍的奈曼(Neyman)的置信区间的概念有所不同. 贝叶斯区间估计其含义有

"先验"和"后验"两种频率解释. 所谓先验解释,就是说,在 θ 的先验分布固定不变的条件下,当多次重复使用同一个贝叶斯区间估计时,事件"$\theta\in[\hat{\theta}_1,\hat{\theta}_2]$"(注意,这里 $\theta,\hat{\theta}_1,\hat{\theta}_2$ 都是随机的),发生的频率大体为 $1-\alpha$. 所谓后验解释,就是说,给定样本值 $x_1,\cdots,x_n$ (这时 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是给定的),当多次使用同一贝叶斯区间估计时,将样本 $X_1=x_1,\cdots,X_n=x_n$ 那些结果挑出来,在这些结果中,事件"$\theta\in[\hat{\theta}_1,\hat{\theta}_2]$"发生的频率大体也为 $1-\alpha$. 在贝叶斯区间估计的两种解释中,θ 始终是随机变量,而区间则有随机或确定两种含义. 而在奈曼的置信区间的概念中,θ 始终是常数,而区间则始终是随机区间. 那么在贝叶斯区间估计中是否容许第三种解释,即 θ 为常数,区间为随机区间呢?回答是否定的. 假定在给定 θ 时,样本的概率分布为 P_θ,区间$[\hat{\theta}_1,\hat{\theta}_2]$满足(12.5)和(12.6)两式,容易举例说明 $P_\theta(\theta\in[\hat{\theta}_1,\hat{\theta}_2])\geqslant 1-\alpha$ 是不成立的. 因此,贝叶斯区间估计与奈曼的置信区间概念之间没有共同点.

此外,贝叶斯区间估计还依赖于 θ 的先验分布,当先验分布不同时,同一水平的区间估计有不同的形式.

例 12.2.1 设 $X_1,\cdots,X_n$ 为抽自 $N(\theta,1)$的简单样本,θ 的先验分布为 $N(0,b^2)$. 求 θ 的水平 $1-\alpha$ 的贝叶斯区间估计.

由例 11.5.14 知,当给定样本 $X_1=x_1,\cdots,X_n=x_n$ 时,θ 的后验分布为 $N\left(\dfrac{nb^2\bar{x}}{1+nb^2},\dfrac{b^2}{1+nb^2}\right)$,其中 $\bar{x}=\dfrac{1}{n}\sum_1^n x_i$. 因此 θ 的一个水平 $1-\alpha$ 的贝叶斯区间估计为

$$\left[\delta^2\overline{X}-z_{\alpha_0}\sqrt{\frac{b^2}{1+nb^2}},\delta^2\overline{X}+z_{\alpha_1}\sqrt{\frac{b^2}{1+nb^2}}\right],$$

其中 $\alpha_0+\alpha_1=\alpha$,z_{α_0}、z_{α_1}分别为 $N(0,1)$的上 α_0 和上 α_1 分位数. $\delta^2=\dfrac{nb^2}{1+nb^2}$. 当 $\alpha_0=\alpha_1=\dfrac{\alpha}{2}$时,区间的中心位置比 $\overline{X}$ 更偏向原点,这显

然是受先验分布 $N(0,b^2)$ 的影响. 另外区间长度 $2z_{\frac{\alpha}{2}}\Big/\sqrt{b_2^{-1}+n}$，比常用的对称置信区间的长度 $\left(2z_{\frac{\alpha}{2}}\Big/\sqrt{n}\right)$ 还小一点儿!

同点估计的情形一样，贝叶斯区间估计的一个主要困难是如何确定一个合理的先验分布. 经验贝叶斯方法似乎可以考虑. 但由于区间估计涉及到计算概率的问题，用经验贝叶斯方法得到的区间估计难免过分粗糙.

13 假设检验

13.1 引言

假设检验(test of hypothesis)是统计推断中的一个重要内容,它是利用搜集到的数据对某个事先作出的统计假设按照某种设计好的方法进行检验,判断此假设是否正确.下面通过一个实例来说明假设检验问题中的基本概念.

例 13.1.1 设某厂生产的一批产品,其次品率 p 是未知的.按规定,若 $p\leqslant 0.01$,则这批产品为可接受的;否则为不可接受的.这里"$p\leqslant 0.01$"便是一个需要检验的假设,记为 H.假定从这批数量很大的产品中随机地抽取 100 件样品,发现其中有三件次品.这一抽样结果便成为判断假设 H 是否成立的依据.显然,样品中次品个数愈多就对假设 H 愈不利;反之则对 H 有利.记样品中次品个数为 X,问题是:X 大到什么程度就应该拒绝 H?由于否定了 H 就等于否定了一大批产品,这个问题应该慎重处理.统计学上的常用的作法是:先假定 H 成立,来计算 $X\geqslant 3$ 的概率有多大,由于 X 的分布为 $B(n,p)$,其中 $n=100$,容易算得 $P_{p=0.01}(X\geqslant 3)\approx 0.08$.显然,对 $p<0.01$,这概率值还要小,也就是说,当假设 $H(p\leqslant 0.01)$成立时,100 个样品中有 3 个或 3 个以上次品的概率不超过 0.08.这可以看作是一个"小概率"事件.而在一次试验中就发生了一个小概率事件是不大可能的.因此,事先作出的假设"$p\leqslant 0.01$"是非常可疑的.在需要作出最终判决时,就应该否定这个假设,而认定这批产品不可接受(即认为 $p>0.01$).

上述例子中包含了假设检验的一些重要的基本概念.一般,设

θ 为用以确定总体分布的一个未知参数，其一切可能值的集合记为 Θ. 则关于 θ 的任一假设可用"$\theta \in \Theta'$"来表示，其中 Θ' 为 Θ 的一个真子集. 在统计假设检验中，首先要有一个作为检验对象的假设，常称为**原假设**或**零假设**(null hypothesis). 与之相应，为使问题表述得更明确，还常提出一个与之对应的假设，称为**备择假设**(alternative hypothesis). 原假设和对立假设常表示为

$$H_0:\theta \in \Theta_0, H_1:\theta \in \Theta_1,$$

其中 Θ_0 和 Θ_1 为 Θ 的两个不相交的真子集，H_0 表示原假设，H_1 表示备择假设.

关于一维实参数的假设常有以下三种形式(其中 θ_0 为给定值)：

(1) 单边假设　$H_0:\theta \leqslant \theta_0, H_1:\theta > \theta_0$.

(2) 单边假设　$H_0:\theta \geqslant \theta_0, H_1:\theta < \theta_0$.

(3) 双边假设　$H_0:\theta = \theta_0, H_1:\theta \neq \theta_0$.

检验假设的依据是样本. 样本的某些取值可能对原假设 H_0 有利，而另一些取值则对 H_0 不利，因此可以根据某种公认的合理准则将样本空间分成两部分. 一部分称为**拒绝域**(critical region)，当样本落入拒绝域时便拒绝 H_0；另一部分可称为**接受域**(acceptance region)，当样本落入它时不拒绝 H_0.

构造拒绝域的常用方法是寻找一个统计量 g(如例 13.1.1 中的样品中次品的件数 X)，g 的大小可以反映对原假设 H_0 有利或不利. 因此，确定拒绝域 W 的问题可转化为确定 g 的一个取值域 C 的问题.

定义 13.1.2　对假设检验问题，设 $X_1, \cdots, X_n$ 为样本，W 为样本空间中的一个子集，对给定的 $\alpha \in (0,1)$，若 W 满足

$$P_\theta((X_1, \cdots, X_n) \in W) \leqslant \alpha, \forall\, \theta \in \Theta_0,$$

则称由 W 构成拒绝域的检验方法为**显著性水平**(evidence level)α 的检验.

以下简称显著性水平为水平. 水平 α 常用的取值为 0.1, 0.05, 0.01 等等. 对一个水平 α 的检验, 假定原假设 H_0 成立, 而样本落入拒绝域 W 中, 就意味着一个小概率事件发生了. 而在一次试验中发生一个小概率事件是可疑的, 结果就导致了对原假设 H_0 的否定. 在例 13.1.1 中, 如果事先给定 $\alpha=0.1$, 而 $P_{p=0.01}(X\geqslant 3)=0.08$, 因此当 $p<0.01$ 时, 这概率还要小. 根据定义 13.1.2 $pW=\{X\geqslant 3\}$ 便给出了假设 $H_0: p\leqslant p_0$ 的一个水平 0.1 的拒绝域. 由 $X=3$ 便可拒绝 H_0. 但如果事先给定水平 $\alpha=0.05$, 这时, 相应的水平 α 的检验有拒绝域 $W=\{X\geqslant 4\}$, 这时 $X=3$ 就不能拒绝 H_0. 由此可见, 水平 α 愈小, 则拒绝原假设愈困难. 换言之, 水平 α 愈小, 则当样本落入拒绝域因而拒绝 H_0 就愈加可信.

通常, 作假设者对原假设 H_0 往往事先有一定的信任度, 或者一旦否定了 H_0 就意味着作出了一个重大的决策, 需谨慎从事, 因此把检验的水平 α 取得比较小其中体现了一种"保护原假设"的思想.

方法 13.1.3 假设检验的一般步骤

(1) 对待检验的未知参数 θ 根据问题的需要作出一个单边或双边的假设. 选择原假设的原则是: 事先有一定信任度或者出于某种考虑是否要加以"保护".

(2) 选定一个显著性水平 α, 最常用的是 $\alpha=0.05$, 放松一点可取 $\alpha=0.075$ 或 0.1, 严格一些可取 $\alpha=0.025$ 或 0.01.

(3) 构造一个统计量 g, g 的大小反映对 H_0 有利或不利. 拒绝域有形式 $W=\{g\in C\}$.

(4) 根据定义 13.1.2 来确定 W.

13.2 正态总体参数的假设检验

13.2.1 单正态总体参数检验

设总体分布为 $N(\mu,\sigma^2)$, $X_1,\cdots,X_n$ 为简单样本, α 为给定的

显著性水平.

方法 13.2.1 方差 σ^2 已知时,均值 μ 的单边检验

$$\text{假设 } H_0: \mu \leqslant \mu_0, H_1: \mu > \mu_0,$$

其中 μ_0 为已知数. 从 μ 的常用估计 $\overline{X} = \frac{1}{n}\sum_1^n X_i$ 出发来构造检验拒绝域. 显然 $\overline{X}$ 愈大则对 H_0 愈不利. 因此拒绝域 W 应有形式 $\{\overline{X} > k\}$,其中临界值 k 应根据定义 13.1.2 由水平 α 来决定. 由于当 $\mu = \mu_0$ 时,统计量 U 为

$$U = \frac{\sqrt{n}\,(\overline{X} - \mu_0)}{\sigma} \sim N(0,1),$$

且 U 为 $\overline{X}$ 的线性函数,用 U 来构造 W 与用 $\overline{X}$ 是等价的. 设 z_α 为 $N(0,1)$的上 α 分位数,有

$$P_{\mu=\mu_0}(U > z_\alpha) = \alpha,$$

显然,当 $\mu < \mu_0$ 时,上述概率小于 α. 因此令 $W = \{U > z_\alpha\} = \left\{\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$作为拒绝域便给出了一个水平 α 的检验.

对相反的单边假设 $H_0: \mu \geqslant \mu_0, H_1: \mu < \mu_0$,相应的水平 α 的检验有拒绝域 $W = \{U < -z_\alpha\} = \left\{\overline{X} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$.

方法 13.2.2 方差 σ^2 已知时,均值 μ 的双边检验

$$\text{假设 } H_0: \mu = \mu_0, H_1: \mu \neq \mu_0,$$

这是一个双边假设. 如同单边假设检验的情形一样,仍然可以从统计量 U 出发来构造检验. 这次是当 $|U|$过大就对原假设 H_0 不利. 设 $z_{\frac{\alpha}{2}}$为 $N(0,1)$的上$\frac{\alpha}{2}$分位数,有

$$P_{\mu=\mu_0}(|U| > z_{\frac{\alpha}{2}}) = \alpha,$$

因此,令

$$W = \{|U| > z_{\frac{\alpha}{2}}\}$$

$$= \left\{\overline{X} < \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\} \cup \left\{\overline{X} + \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\},$$

则 W 即为水平 α 的检验的拒绝域.

方法 13.2.3 方差 σ^2 未知时，均值 μ 的单边检验

$$\text{假设 } H_0: \mu \leqslant \mu_0, H_1: \mu > \mu_0.$$

我们仍然从 $\overline{X}$ 出发来构造检验的拒绝域，但统计量 U 已不能用，因为 U 是未知参数 σ 的函数. 如同在区间估计的场合，在统计量 U 中用样本标准差 $S=\sqrt{\frac{1}{n-1}\sum\limits_{1}^{n}(X_i-\overline{X})^2}$ 来代替总体标准差 σ，便得到

$$T = \sqrt{n}\,(\overline{X} - \mu_0)/S.$$

当 $\mu=\mu_0$ 时，$T \sim t(n-1)$（见定理 6.4.21）. 由于 S 的分布与 μ 无关，且 S 与 $\overline{X}$ 独立，仍然可以将 T 作为 $\overline{X}$ 的线性函数看待. 因此当 T 的值愈大时对原假设 H_0 愈不利. 水平 α 的检验有拒绝域

$$W = \{T > t_\alpha(n-1)\} = \left\{\overline{X} > \mu_0 + t_\alpha(n-1)\frac{S}{\sqrt{n}}\right\}.$$

其中 $t_\alpha(n-1)$ 为 $t(n-1)$ 的上 α 分位数.

对相反方向的单边假设 $H_0: \mu \geqslant \mu_0, H_1: \mu < \mu_0$，相应水平的检验有拒绝域 $W=\{T<-t_\alpha(n-1)\}=\left\{\overline{X}<\mu_0-t_\alpha(n-1)\frac{S}{\sqrt{n}}\right\}$.

方法 13.2.4 方差 σ^2 未知时，均值 μ 的双边检验

$$\text{假设 } H_0: \mu = \mu_0, H_1: \mu \neq \mu_0.$$

这时 $|T|$ 愈大则对 H_0 愈不利. 因此水平 α 的检验有拒绝域

$$\begin{aligned} W &= \{|T| > t_{\frac{\alpha}{2}}(n-1)\} \\ &= \left\{\overline{X} < \mu_0 - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right\} \\ &\quad \cup \left\{\overline{X} > \mu_0 + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right\}, \end{aligned}$$

其中 $t_{\frac{\alpha}{2}}(n-1)$ 为 $t(n-1)$ 的上 $\frac{\alpha}{2}$ 分位数.

方法 13.2.5 均值 μ 已知时,方差 σ^2 的检验

$$\text{假设 } H_0: \sigma^2 \leqslant \sigma_0^2, H_1: \sigma^2 > \sigma_0^2.$$

其中 σ_0^2 为已知数. 我们从 σ^2 的常用估计 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i}^{n}(X_i - \mu)^2$ 出发来构造检验的拒绝域. 显然,当 $\hat{\sigma}^2$ 愈大对原假设 H_0 就愈不利. 因此拒绝域应有 $\{\hat{\sigma}^2 > C\}$ 的形式. 当 $\sigma^2 = \sigma_0^2$ 时,

$$K = n\hat{\sigma}^2/\sigma_0^2 \sim \chi^2(n).$$

因此水平 α 的检验有拒绝域

$$W = \{K > \chi_\alpha^2(n)\} = \left\{\hat{\sigma}^2 > \sigma_0^2 \frac{\chi_\alpha^2(n)}{n}\right\},$$

其中 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 的上 α 分位数.

对相反方向的单边假设 $H_0: \sigma^2 \geqslant \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$,相应水平的检验有拒绝域

$$W = \{K < \chi_{1-\alpha}^2(n)\} = \left\{\hat{\sigma}^2 < \sigma_0^2 \frac{\chi_\alpha^2(n)}{n}\right\},$$

其中 $\chi_{1-\alpha}^2(n)$ 为 $\chi^2(n)$ 的上 $1-\alpha$ 分位数.

对双边假设 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$,水平 α 的检验有拒绝域

$$\begin{aligned} W &= \{K < \chi_{1-\alpha_0}^2(n)\} \cup \{K > \chi_{\alpha_1}^2(n)\} \\ &= \left\{\hat{\sigma}^2 < \sigma_0^2 \frac{\chi_{1-\alpha_0}^2(n)}{n}\right\} \cup \{\hat{\sigma}^2 > \sigma_0^2\chi_{\alpha_1}^2(n)/n\}, \end{aligned}$$

其中 $\alpha_0 + \alpha_1 = \alpha$,实用中常取 $\alpha_0 + \alpha_1 = \alpha$.

方法 13.2.6 均值 μ 未知时,方差 σ^2 的检验

$$\text{假设 } H_0: \sigma^2 \leqslant \sigma_0^2, H_1: \sigma^2 > \sigma_0^2.$$

由于 μ 未知,我们从 σ^2 的常用估计 $S^2 = \frac{1}{n-1}\sum_{1}^{n}(X_i - \overline{X})^2$ 出发来构造检验的拒绝域. 当 $\mu = \mu_0$ 时,$K = (n-1)S^2/\sigma_0^2 \sim \chi^2(n-1)$. 因此,对这种单边假设,水平 α 的检验有拒绝域

$$W=\{K>\chi_\alpha^2(n-1)\}=\{S^2>\sigma_0^2\chi_\alpha^2(n-1)/(n-1)\}.$$

对相反的假设 $H_0:\sigma^2\geqslant\sigma_0^2, H_1:\sigma^2<\sigma_0^2$，相应水平的检验有拒绝域

$$W=\{K<\chi_{1-\alpha}^2(n-1)\}=\{S^2<\sigma_0^2\chi_{1-\alpha}^2(n-1)/(n-1)\}.$$

对双边假设 $H_0:\sigma^2=\sigma_0^2, H_1:\sigma^2\neq\sigma_0^2$，水平 α 的检验有拒绝域

$$\begin{aligned}W&=\{K<\chi_{1-\alpha_0}^2(n-1)\}\cup\{K>\chi_{\alpha_1}^2(n-1)\}\\&=\{S^2<\sigma_0^2\chi_{1-\alpha_0}^2(n-1)/(n-1)\}\\&\quad\cup\{S^2>\sigma_0^2\chi_{\alpha_1}^2(n-1)/(n-1)\},\end{aligned}$$

其中 $\alpha_0+\alpha_1=1$.

13.2.2 双正态总体参数检验

设 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$ 为分别取自总体 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$ 的简单样本，记 $\delta=\mu_1-\mu_2, \lambda=\sigma_1^2/\sigma_2^2$，$\delta$ 和 λ 的假设检验比区间估计有更广泛的应用.

方法 13.2.7 方差 σ_1^2 和 σ_2^2 均已知时，δ 的检验

对下述三种假设

(1) $H_0:\delta\leqslant\delta_0, H_1:\delta>\delta_0$,

(2) $H_0:\delta\geqslant\delta_0, H_1:\delta<\delta_0$,

及

(3) $H_0:\delta=\delta_0, H_1:\delta\neq\delta_0$.

用于构造检验的统计量均可采用

$$U=(\overline{X}-\overline{Y}-\delta_0)\Big/\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}.$$

当 $\delta=\delta_0$ 时 $U\sim N(0,1)$. 因此对假设(1)(2)及(3)，检验拒绝域分别为

$$W_1=\{U>z_\alpha\},$$

$$W_2=\{U<-z_\alpha\},$$

及

$$W_3=\{|U|>z_{\frac{\alpha}{2}}\},$$

其中 z_α 及 $z_{\frac{\alpha}{2}}$ 分别为 $N(0,1)$ 的上 α 及上 $\frac{\alpha}{2}$ 分位数.

方法 13.2.8 $\sigma_1^2=\sigma_2^2$ 但未知时,δ 的检验

对 13.2.7 中的三个假设,用于构造检验的统计量统一可用

$$T=\sqrt{\frac{nm}{n+m}}(\overline{X}-\overline{Y}-\delta_0)/S_p,$$

其中 $S_p=\left[\frac{\sum_1^n(X_i-\overline{X})^2+\sum_1^m(Y_j-\overline{Y})^2}{n+m-2}\right]^{\frac{1}{2}}$. 当 $\delta=\delta_0$ 时,$T\sim t(n+m-2)$. 因此对 13.2.7 中的假设(1),(2)及(3),检验拒绝域分别为

$$W_1=\{T>t_\alpha(n+m-2)\},$$

$$W_2=\{T<-t_\alpha(n+m-2)\},$$

及

$$W_3=\{|T|>t_{\frac{\alpha}{2}}(n+m-2)\},$$

其中 $t_\alpha(n+m-2)$ 及 $t_{\frac{\alpha}{2}}(n+m-2)$ 分别为 $t(n+m-2)$ 的上 α 及上 $\frac{\alpha}{2}$ 分位数.

方法 13.2.9 σ_1^2 与 σ_2^2 均未知且不相同时,δ 的检验

这时检验统计量可用(12.1)中定义的 T'(其中 δ 用 δ_0 代替),当 $\delta=\delta_0$ 时,$T'\sim t(n-1)$. 因此对方法 13.2.7 中的三个假设,检验拒绝域分别为

$$W_1=\{T'>t_\alpha(n-1)\},$$

$$W_2=\{T'<-t_\alpha(n-1)\},$$

$$W_3=\{|T'|>t_{\frac{\alpha}{2}}(n-1)\}.$$

另外,检验统计量又可用定理 12.1.14 中的 T''(同样,其中 δ 用 δ_0 代入),则当 $\delta=\delta_0$ 时,$T''\sim t(\nu)$,ν 也由该定理给出. 这时方法 13.2.7 中的三个假设分别有检验拒绝域

$$W_1 = \{T'' > t_\alpha(\nu)\},$$

$$W_2 = \{T'' < -t_\alpha(\nu)\},$$

及

$$W_3 = \{|T''| > t_{\frac{\alpha}{2}}(\nu)\}.$$

方法 13.2.10 均值 μ_1,μ_2 均已知时，方差比 λ 的检验

对下述三种假定

(1) $H_0: \lambda \leqslant \lambda_0, H_1: \lambda > \lambda_0$,

(2) $H_0: \lambda \geqslant \lambda_0, H_1: \lambda < \lambda_0$,

及 (3) $H_0: \lambda = \lambda_0, H_1: \lambda \neq \lambda_0$.

均用检验统计量

$$F = \frac{1}{\lambda_0} \frac{\frac{1}{n}\sum_1^n (X_i - \mu_1)^2}{\frac{1}{m}\sum_1^m (Y_j - \mu_2)^2}.$$

当 $\lambda = \lambda_0$ 时，$F \sim F(n,m)$. 因此对方法 13.2.10 中的三个假设，检验的拒绝域分别为

$$W_1 = \{F > F_\alpha(n,m)\},$$

$$W_2 = \{F < F_{1-\alpha}(n,m)\},$$

及

$$W_3 = \{F < F_{1-\alpha_0}(n,m)\} \cup \{F > F_{\alpha_1}(n,m)\},$$

其中 $\alpha_0 + \alpha_1 = \alpha$, $F_\alpha(n,m)$等均为 $F(n,m)$的上分位数.

方法 13.2.11 均值 μ_1,μ_2 均未知时，λ 的检验

这时，对方法 13.2.10 中的假设(1),(2)及(3),统一用检验统计量

$$F = \frac{1}{\alpha_0} \frac{S_X^2}{S_Y^2},$$

其中 $S_X^2 = \frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2, S_Y^2 = \frac{1}{m-1}\sum_1^m (Y_j - \overline{Y})^2$. 当 $\lambda = \lambda_0$ 时，$F \sim F(n-1, m-1)$. 因此对方法 13.2.10 中的三个假设，检

验的拒绝域分别为

$$W_1 = \{F > F_\alpha(n-1, m-1)\},$$

$$W_2 = \{F < F_{1-\alpha}(n-1, m-1)\},$$

及

$$W_3 = \{F < F_{1-\alpha_0}(n-1, m-1)\} \cup \{F > F_{\alpha_1}(n-1, m-1)\},$$

其中 $\alpha_0+\alpha_1=\alpha$, F_α 等均为 $F(n-1,m-1)$ 的上分位数.

例 13.2.12 某工厂内有两台机床生产滚珠. 为比较这两台机床的性质,从它们生产的产品中分别抽出 8 个和 9 个,测得滚珠的直径与设计值的偏差列于下表.

表 13.1 单位:μm

机床 A	0, −5, 2, 5, −2, 1, 2, −2
机床 B	2, 0, −2, 2, 0, 0, −2, 1, −2

这里涉及到两个假设检验问题. 首先要检验两台机床各自的偏差均值是否为 0? 其次要比较两台机床的加工精度,即比较偏差方差的大小. 用 $x_1,\cdots,x_8$ 表示机床 A 的样本, $y_1,\cdots,y_9$ 表示机床 B 的样本. 可算得 $\bar{x}=0.125$, $S_x=3.091$, $\bar{y}=-0.111$, $S_y=1.616$. 由于方差未知,我们用方法 13.2.3 给出的 t 检验来检验均值. 算得 $T_x=0.114$, $T_y=-0.206$. 设检验水平为 $\alpha=0.1$, 由于要检验均值是否为零,这是个双边检验问题. 通过查 t 分布表(附表 3),有 $t_{0.05}(7)=1.8946$, 而 $t_{0.05}(8)=1.8595$. 显然有 $|T_x|<t_{0.05}(7)$, $|T_y|<t_{0.05}$. 因此可以推定这二台机床的偏差均值都为零. 下一个检验是方差比. 由于 $S_y<S_x$, 我们自然可假定 B 机床的方差比 A 机床小. 因此相当于检验假设 13.2.10(2),其中 $\lambda_0=1$. 相应的检验统计量为 $F=S_x^2/S_y^2=3.659$, 通过查 F 分布表(附表 4),有 $F_{0.1}(7,8)=2.62$. 显然 $F>F_{0.1}(7,8)$. 因此可以断定 A 机床的方差比 B 机床的大,也就是说, B 机床的加工精度比 A 机床的高.

13.3 两种错误概率及检验的评优准则

13.3.1 两种错误概率

对假设检验问题 13.1.1,设某个检验的拒绝域为 W. 当检验者根据样本值落在 W 内与否而拒绝或保留原假设时,可能会犯两种错误.

定义 13.3.1 (1) 当原假设 $H_0:\theta\in\Theta_0$ 成立,而样本值却落在拒绝域 W 内(因而要拒绝 H_0)时,称为犯**第一类错误**(error of the first kind). 犯第一类错误的概率为 $P_\theta(W),\theta\in\Theta_0$.

(2) 当备择假设 $H_1:\theta\in\Theta_1$ 成立,而样本值却落在 W 外(因而要保留 H_0)时,称为犯**第二类错误**(error of the second kind). 犯第二类错误的概率为 $P_\theta(\overline{W}),\theta\in\Theta_1$.

第一类错误又称为**弃真错误**;第二类错误又称为**存伪错误**. 设两个检验分别有拒绝域 W 和 W',$W\subset W'$,显然前一检验犯第一类错误的概率比后一检验小,而犯第二类错误的概率大. 而一个水平 α 的检验满足条件:犯第一类错误的概率不超过 α. 由于通常水平 α 取得比较小,这就体现了"保护原假设"或者说"限制犯第一类错误概率"的原则.

定义 13.3.2 对一个具有拒绝域 W 的检验,其**功效函数**(power function)为

$$\beta(\theta)=P_\theta(W),\theta\in\Theta,$$

$\beta(\theta)$又简称为**功效**.

当 $\theta\in\Theta_1$(备择假设)时,$1-\beta(\theta)$即为犯第二类错误的概率,因此当 $\theta\in\Theta_1$ 时,要求 $\beta(\theta)$愈大愈好.

13.3.2 一致最优检验

定义 13.3.3 对给定的 $\alpha\in(0,1)$,设 τ 为一个水平 α 的检

验. 若对任一水平 α 的检验 τ' 有

$$\beta_\tau(\theta) \geqslant \beta_{\tau'}(\theta), \forall\ \theta \in \Theta_1,$$

其中 $\beta_\tau(\theta)$ 和 $\beta_{\tau'}(\theta)$ 分别为 τ 和 τ' 的功效函数, 则称 τ 为**一致最优检验**(uniformly most powerfwl test).

一致最优检验简称为 UMP 检验. 以下用 $X=(X_1,\cdots,X_n)$ 和 $x=(x_1,\cdots,x_n)$ 分别记样本和样本值, 用 $L(X,\theta)$ 记似然函数.

引理 13.3.4 奈曼-皮尔逊(Neymann-Pearson)**引理** 对下述的简单假设

$$H_0:\theta = \theta_0, H_1:\theta = \theta_1,$$

对给定的 $\alpha \in (0,1)$, 一定存在水平 α 的 UMP 检验 τ, 它的拒绝域有形式

$$W = \{x: 当\ L(x,\theta_1)/L(x,\theta_0) > C_\alpha\},$$

其中 C_α 为由 α 决定的常数.

注 引理 13.3.4 是针对原假设和备择假设都只有一个点时得到的结论. 当样本分布连续时, 一定可找到一个 C_α, 使得 $P_{\theta_0}(W)=\alpha$. 而当样本分布为离散型时, 就作不到这点. 这时 C_α 应取得使 $P_{\theta_0}(W)$ 小于 α 但尽可能靠近 α.

由引理 13.3.4 可以推出一个重要的结论, 即对相当广泛的一类分布族, 单边假设检验问题的 UMP 检验是存在的.

定理 13.3.5 设 θ 为一维未知参数, $L(X,\theta)$ 为似然函数, 满足: 对任意给定的 $\theta_1<\theta_2$, 比值

$$L(X,\theta_2)/L(X,\theta_1) = g(T(X),\theta_1,\theta_2).$$

其中 g 作为样本 X 的函数只与 $T(X)$ 有关, 且关于 $T(X)$ 单调非降, 则对单边假设

$$H_0:\theta \leqslant \theta_0,\ H_1:\theta > \theta_0$$

及给定的 $\alpha\in(0,1)$，水平 α 的 UMP 检验存在，且其拒绝域有形式

$$W = \{x:\text{当 } T(x) > C_\alpha\}.$$

例 13.3.6 设总体分布为 $N(\mu,\sigma_0^2)$，其中 μ 未知，σ_0^2 已知，$X_1,\cdots,X_n$ 为简单样本. 于是对 $\mu_1<\mu_2$，似然比

$$L(x,\mu_2)/L(x,\mu_1) = \exp\left\{\frac{\mu_2-\mu_1}{n\sigma_0^2}\left(\frac{1}{n}\sum_1^n X_i\right)\right\},$$

因此似然比为 $\overline{X} = \frac{1}{n}\sum_1^n X_i$ 的单调函数. 由定理 13.3.5 知，对单边假设

$$H_0:\mu\leqslant\mu_0, H_1:\mu>\mu_0.$$

水平 α 的 UMP 检验有拒绝域

$$W = \{\overline{X} > C_\alpha\}.$$

根据 13.2.2 知 $C_\alpha=\mu_0+z_\alpha\left(\frac{\sigma_0}{\sqrt{n}}\right)$.

13.3.3 无偏检验

双边检验问题一般不存在 UMP 检验. 为此，类似在点估计问题中引进无偏估计的概念，我们引入无偏检验的概念.

定义 13.3.7 设对假设 $H_0:\theta\in\Theta_0, H_1:\theta\in\Theta_1$，$\tau$ 为一检验，其功效为 $\beta(\theta)$. 若对任意 $\theta_0\in\Theta_0$ 及 $\theta_1\in\Theta_1$ 有 $\beta(\theta_0)\leqslant\beta(\theta_1)$，则称 τ 为一个**无偏检验**(unbiased test).

由于当 $\theta\in\Theta_0$ 时 $\beta(\theta)$ 为犯第一类错误的概率，而当 $\theta_0\in\Theta_1$ 时，$1-\beta(\theta)$ 为犯第二类错误的概率，定义 13.3.8 无异是要求一个检验满足犯弃真错误的概率总不超过存伪的概率.

容易验证方法 13.2.2 和 13.2.4 中给出的检验均为无偏检验. 在定义 13.3.3 中，若将涉及的检验都限定为无偏检验，便可得到一致最优无偏检验的概念. 关于这方面的深入讨论请参考文献[13].

13.4 似然比检验与渐近水平检验

奈曼和皮尔逊根据似然原理和由他们给出的引理 13.3.4 给出了一种构造检验的一般方法.

定义 13.4.1 设 θ 为一个 k 维未知参数,$L(X,\theta)$ 为似然函数. 对假设 $H_0:\theta\in\Theta_0, H_1:\theta\in\Theta_1$,规定**似然比**(likelihood ratio)为

$$LR(x)=\sup_{\theta\in\Theta_0\cup\Theta_1}L(x,\theta)/\sup_{\theta\in\Theta_0}L(x,\theta),$$

具有拒绝域 $W=\{x:LR(x)>c\}$ 的检验称为**似然比检验**.

在似然比中,分子部分代表给定样本时 θ 的"全局似然性",而分母部分代表关于原假设的"局部似然性". 因此似然比检验中含意为:当关于原假设的局部似然性相当小时(与全局似然性相比),就要拒绝 H_0. 似然比检验中的一个关键问题是要计算似然比 LR 的分布. 在双边假设检验问题中,即使是在有多余参数的场合,关于似然比 LR 有一个渐近分布.

设 k 维未知参数 $\theta=(\theta_1,\theta_2)$,其中 θ_1 为 d 维($d<k$),θ_2 为 $k-d$ 维. 考虑假设

$$H_0:\theta_1=\theta_{10}, H_1:\theta_1\neq\theta_{10}, \tag{13.1}$$

其中 θ_{10} 为给定 d 维常向量. 注意此假设与 θ_2 无关. 则对许多常见的分布族有下述结论:对简单样本 $X_1,X_2,\cdots,X_n$,

$$\lim_{n\to\infty}2\log(LR)\xlongequal{\mathrm{d}}\chi^2(k-d). \tag{13.2}$$

定理 13.4.2 设总体分布族的频率函数有以下形式:

$$f(x,\theta)=C(\theta)\exp\Big\{\sum_{i=1}^{k}T_i(x)g_i(\theta)\Big\}h(x),$$

其中$(g_1(\theta),\cdots,g_k(\theta))$为与 θ 一一对应的变换,则(13.2)的结论成立.

在计算似然比 LR 时,它的分子和分母有时不能直接求得,这

时可以用似然方程的解代入，而渐近分布仍然成立. 利用式(13.2)的结论，可以给出假设式(13.1)的渐近水平为 α 的似然比检验具有拒绝域

$$W = \{x: 2\log LR(x) > \chi_\alpha^2(k-d)\}. \tag{13.3}$$

除了利用似然比的渐似分布之外，还可利用中心极限定理给出渐近水平检验，这样做的优点是对单边假设检验也能行得通.

例 13.4.3 设 $X_1, \cdots, X_n$ 为简单样本，$P(X_1=1)=p, P(X_1=0)=1-p$，其中 p 为未知参数. 则由中心极限定理，当 $p=p_0$ 时，

$$\frac{\sqrt{n}(\overline{X}-p_0)}{\sqrt{p_0(1-p_0)}} \xrightarrow{\mathrm{d}} N(0,1), \text{当 } n \to \infty,$$

其中 $\overline{X} = \frac{1}{n}\sum_1^n X_i$. 因此，对单边假设

$$H_0: p \leqslant p_0, H_1: p > p_0,$$

可给出一个渐近水平 α 的检验，其拒绝域为 $W = \left\{x: \frac{\sqrt{n}(\overline{X}-p_0)}{\sqrt{p_0(1-p_0)}} > z_\alpha\right\}$，其中 z_α 为 $N(0,1)$ 的上 α 分位数.

13.5 拟合优度检验

在前几节中介绍的假设检验问题统称为参数检验问题，即事先认为样本分布具有某种指定的形式，而其中有一些未知的参数，检验的目标是关于某个参数落在特定范围内的假设. 本节要介绍的是另一类假设检验，其目标不是针对具体的参数，而是针对分布的类型. 例如，通常总假定总体分布具有正态性，则“总体分布为正态”这一断言本身在一定场合下就是可疑的，有待于检验. 又如，有两个随机变量 X 与 Y，有时要在不假定 X 与 Y 具有某种分布类型

时检验它们之间的独立性. 一般的,可将这类假设表述如下:设$\{F_\theta,\theta\in\Theta\}$为一个分布族,则提出假设

$$H:\text{总体 } X \text{ 的分布 } F \in \{F_\theta,\theta \in \Theta\} \tag{13.4}$$

等待检验. 在这假设中只关心 F 是否属于$\{F_\theta,\theta\in\Theta\}$的问题,而不计较 θ 究竟取何值. 检验假设 H 的基本方法是用样本 $X_1,\cdots,X_n$ 去拟合$\{F_\theta,\theta\in\Theta\}$中的一个分布,根据拟合的好坏来决定对 H 的取舍.

皮尔逊(K Pearson)早在 1900 年就提出了一个用于检验假设(13.4)的统计量,这就是著名的 **χ^2 统计量**. 为说明 χ^2 统计量的构造与应用,先假定分布族$\{F_\theta,\theta\in\Theta\}$中只有一个分布 F_0,这时(13.4)转化为

$$H:F = F_0. \tag{13.5}$$

总体分布不一定是一维的. 这时,可将总体取值的范围分成 m 个两两不相交的子集 $S_1,S_2,\cdots,S_m$. 当总体为一维时,$S_1,\cdots,S_m$ 即为 m 个互不相交的区间,记

$$p_i = P_{F_0}(X \in S_i),i = 1,\cdots,m.$$

对简单样本 $X_1,\cdots,X_n$,记

$$n_i = X_1,\cdots,X_n \text{ 落在 } S_i \text{ 内的个数},i = 1,\cdots,m.$$

则$\left(\frac{n_i}{n}-p_i\right)^2$表示样本落在 S_i 内的频率关于 S_i 的概率的偏差. 皮尔逊用$\left(\frac{n_i}{n}-p_i\right)^2$加权和

定义 13.5.1

$$K_n = \sum_{i=1}^{m} n\left(\frac{n_i}{n} - p_i\right)^2\Big/p_i = \sum_{i=1}^{m}(n_i - np_i)^2/np_i$$

来刻画样本 $X_1,\cdots,X_n$ 拟合分布 F_0 的优度,K_n 称为**皮尔逊 χ^2 统计量**.

定理 13.5.2 当假设(13.5)成立时,皮尔逊 χ^2 统计量 K_n 当 $n\to\infty$时有极限分布 $\chi^2(m-1)$.

显然，当 K_n 较小时表示样本对 F_0 拟合得好，当 K_n 较大时表示这种拟合不太好. 因此对给定水平 α，构造一个检验：当 $K_n > \chi_\alpha^2(m-1)$ 时就否定 H，否则就接受 H. 这就是一个渐近水平 α 的检验.

下面考虑 $\{F_\theta, \theta\in\Theta\}$，其中 θ 为 k 维参数，Θ 为 R^k 中的一个非空区域. 这时在定义 13.5.1 中，K_n 的表达式里的 p_i 与 θ 有关，因为

$$p_i(\theta) = P_{F_\theta}(X \in S_i) \quad i = 1,\cdots,m, \theta \in \Theta. \tag{13.6}$$

这时可用样本构造 θ 的一个估计 $\hat{\theta}_n$，在 $p_i(\theta)$ 中用 $\hat{\theta}_n$ 代替 θ，再代入 K_n 中，得

$$\begin{aligned}\hat{K}_n &= \sum_{i=1}^{m} n\left(\frac{n_i}{n} - p_i(\hat{\theta}_n)\right)^2 \Big/ p_i(\hat{\theta}_n) \\ &= \sum_{1}^{m} (n_i - p_i(\hat{\theta}_n))^2 / np_i(\hat{\theta}_n).\end{aligned} \tag{13.7}$$

用 $\hat{K}_n$ 来作**拟合优度**(goodness of fit). 皮尔逊建议用 $\hat{\theta}_n$ 为下述方程组的解：

$$\sum_{i=1}^{m} \frac{n_i}{np_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0, \quad j = 1,\cdots,k. \tag{13.8}$$

定理 13.5.3 设 $\dfrac{\partial p_i(\theta)}{\partial \theta_j}$ 关于 θ 连续，$i=1,\cdots,m, j=1,\cdots,k$，且 $\hat{\theta}_n$ 为方程组(13.8)的相合解($\hat{\theta}_n$ 为 θ 的相合估计)，则当 $m > k+1$ 时有

$$\hat{K}_n \xrightarrow{\mathrm{d}} \chi^2(m-k-1), \text{当 } n \to \infty.$$

注 总体取值域的划分 $S_1,\cdots,S_m$ 有一定任意性，它与 m 的大小及 $S_1,\cdots,S_m$ 的形状有关. 在实际应用时，m 的大小应与样本 n 相联系，当 n 较大时相应的 m 可较大；反之，n 小时 m 应小. 另外，有时计算(13.8)的解比较麻烦，为了简便可计算 θ 的关于分布族 $\{F_\theta, \theta\in\Theta'\}$ 的极大似然估计来代入 K_n. 这时定理 13.5.3 的结

论可能不成立，但相差不大.

用皮尔逊 χ^2 统计量可检验样本是否取自正态总体 $N(\mu,\sigma^2)$（这里 μ 和 σ^2 可以是任意的）或是否取自皮尔逊总体 $P(\lambda)$（这里 λ 也可以是任意的），等等. 除此之外还有一些重要的应用.

例 13.5.4 独立性检验 设两个随机变量 X,Y 均为离散型的，X 取值于 $\{x_1,\cdots,x_I\}$，Y 取值于 $\{y_1,\cdots,y_J\}$. 设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 为简单样本，记 n_{ij} 为 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 中等于 (x_i,y_j) 的个数，要据此检验假设 H：X 与 Y 独立.

在这种问题中，常把数据列为表 13.1 的形式，称为**列联表**(contingency table)：

表 13.1 列联表

Y \ X	y_1	$\cdots$	y_j	$\cdots$	y_J	
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_I	n_{I1}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\cdot}$
	$n_{\cdot 1}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot J}$	

在此表中，最右一列 $n_{i\cdot}=\sum_{j=1}^{J}n_{ij}$，为第 i 行之和，$i=1,\cdots,I$；最下一行 $n_{\cdot j}=\sum_{i=1}^{I}n_{ij}$，为第 j 列之和，$j=1,\cdots,J$.

记 $p_{ij}=P(X=x_i,Y=y_j)$，$p_{i\cdot}=P(X=x_i)=\sum_{j=1}^{J}p_{ij}$，$p_{\cdot j}=P(Y=y_j)=\sum_{i=1}^{I}p_{ij}$. 则假设 H 可表为

$$H: p_{ij} = p_{i\cdot}\, p_{\cdot j}, i = 1, \cdots, I, j = 1, \cdots, J. \qquad (13.9)$$

将(X,Y)的每一组可能取值作为一个$S_l, l=1,\cdots,m$，则$m=IJ$.而在假设H中有$I+J-2$个自由参数(因有两个约束条件：$\sum_{i=1}^{I} p_{i\cdot} = 1, \sum_{j=1}^{J} p_{\cdot j} = 1$). 用$p_{1\cdot},\cdots,p_{I-1\cdot}$及$p_{\cdot 1},\cdots,p_{\cdot J-1}$作为未知参数代入方程组 13.8 中可解得

$$\hat{p}_{i\cdot} = n_{i\cdot}/n, \qquad i = 1, \cdots, I,$$
$$\hat{p}_{\cdot j} = n_{\cdot j}/n, \qquad j = 1, \cdots, J.$$

将这组解代入$\hat{K}_n$中有

$$\hat{K}_n = n \sum_{i=1}^{I} \sum_{j=1}^{J} (n_{ij} - n_{i\cdot}\, n_{\cdot j}/n)^2 / n_{i\cdot}\, n_{\cdot j}$$
$$= n\left(\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{n_{ij}^2}{n_{i\cdot}\, n_{\cdot j}} - 1 \right).$$

由定理 13.5.3 知$\hat{K}_n$有极限分布$\chi^2(m-k-1)$，$m=IJ$，$k=I+J-2$，因此$m-k-1=IJ-I-J+1=(I-1)(J-1)$. 因此对给定水平$\alpha$，当$\hat{K}_n>\chi_\alpha^2((I-1)(J-1))$时就否定$H$，即断定$X$与$Y$是相关的. 当$n$很大时，这检验的真实水平接近$\alpha$.

除了皮尔逊的χ^2检验之外，还有用其它方法作的拟合优度检验. 科尔莫戈罗夫(Колмогоров)对简单假设(13.5)给出一个检验方法：设$X_1,\cdots,X_n$为简单样本，记$X_{(1)},\cdots,X_{(n)}$为由样本生成的顺序统计量. 记$d_k = \left| \frac{k}{n} - F_0(X_{(k)}) \right|$，$d_k' = \left| \frac{k-1}{n} - F_0(X_k) \right|$，$k=1,\cdots,n$. 定义

$$D_n = \max\{d_k, d_k', k = 1, \cdots, n\}.$$

定理 13.5.5 当假设 13.5.2 成立时

$$\lim_{n\to\infty} P(\sqrt{n}\, D_n < \lambda) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2), \forall\, \lambda > 0.$$

利用此定理可得到假设(13.5)的一个近似检验. 记定理中等式右边为$Q(\lambda)$，对任给$\alpha\in(0,1)$找λ_α使得$1-Q(\lambda_\alpha)=\alpha$，则当$\sqrt{n}\, D_n$

$>\lambda_a$ 时否定 H. 这检验称为科尔莫戈罗夫检验，它比皮尔逊的 χ^2 检验灵敏(所需样本容量小)，但缺点是不能用于复杂假设(13.4)的情形.

D_n 的常用的理论表达形式为

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)|,$$

其中 $F_n(x)$ 为由样本生成的经验分布函数.

例 13.5.6 为考察吸烟与慢性气管炎的关系，调查了 340 名 45 岁至 50 岁之间的人，得下述列联表 13.2.

表 13.2 单位：人

X \ Y	0	1	$\sum$
0	122	13	135
1	162	43	205
$\sum$	284	56	340

其中 $X=0$ 表示不吸烟，$X=1$ 表示吸烟；$Y=0$ 表示未患慢性气管炎，$Y=1$ 表示患慢性气管炎. 如果假设吸烟与患慢性气管炎无关，相当于假设变量 X 与 Y 独立. 这是一个独立性检验问题. 用皮尔逊 χ^2 统计量

$$\hat{K}_n = 7.616.$$

通过查表，$\hat{K}_n$ 介于 $\chi^2_{0.01}(1)(=6.635)$ 与 $\chi^2_{0.005}(1)(=7.879)$ 之间. 由于 0.01 是相当小的概率. 因此，由 $\hat{K}_n>\chi^2_{0.01}(1)$ 可推翻原假设，因此断定吸烟与慢性气管炎有关. 由表 13.2 可以看出，在不吸烟者中，患慢性气管炎的比例为 13/135=9.6%. 而在吸烟者中患慢性气管炎的占 43/205=21.0%. 这说明吸烟可导致患慢性气管炎的机会增大.

14 非参数统计

14.1 引言

在第 11 至第 13 章中，讨论统计推断问题时，总假设样本来自具有已知分布形式的总体，其中有若干未知参数. 用这样的统计模型来讨论问题都属于**参数统计**(parametric statistics)的范畴. 但在实际中也常会遇到下列两种不同的情况.

首先，在许多实用的领域内处理数据时经常无法事先断定样本分布具有某种具体的数学形式，其原因可能是对所研究的领域缺乏经验，或者没有一种客观的背景去推测样本分布应具有的形式. 也就是说不能建立一个由有限个参数确定的统计模型. 当然，为了数学推论的需要，对统计模型总要作一些必要而合理的假定，例如假定总体分布有密度函数，样本为独立同分布的，等等. 在这种很一般的前提下提出来的统计推断问题都属于**非参数统计**(nonparametric statistics)的范畴. 在非参数统计中提问题的方式往往与参数统计不同. 例如假定我们手头有采自两个总体的一批样本. 通过对样本进行粗略的观察，怀疑这两个总体的分布不同. 在参数统计中可以假定两个总体都是正态的. 这时比较两个总体的异同的问题就可以转化为比较各自的均值和方差的问题. 而在非参数统计中就只能直接比较两个总体的分布函数是否相同. 另外，在相同的统计问题中，参数统计与非参数统计所使用的方法也不同. 因为在参数统计中使用的方法都与特定的分布形式有紧密联系，当分布形式发生变化时，原来效果很好的统计方法就会变坏. 例如，用样本均值 $\bar{x}$ 去估计正态分布的对称中心时效果很好，

此时 $\overline{X}$ 是最小方差无偏估计；但用 $\overline{X}$ 去估计柯西(Cauchy)分布的对称中心时则效果很差. 因为它的均值根本不存在，且均方误差为无穷大. 而在非参数统计中建议使用样本中位数 m 去估计总体分布的对称中心，适用于相当大一类总体分布(包括正态和柯西这样性质差别很大的分布).

其次，在搜集数据的过程中，常常由于某些客观原因而使得数据受到"污染"，因而造成统计模型与实际情况发生偏离. 例如，在正态样本中混杂了一些非正态的样本点或者由于观测过程的失误而产生了一些**离群值**(outlier). 当然，对上述情况的发生与否，试验者不能完全断定，而仅仅是一种怀疑. 解决这类问题的方法有两条途径可循. 一是用参数统计的方法对数据进行"**诊断**"(diagnostic)，以判断某些可疑的数据是"好"的，还是"坏"的. 一是寻找一些新的统计方法或准则，使得当模型假定正确时，得出的结论有接近最优的性质；而当模型发生偏差时，得出的结论仍具有较好的性质，或至少不能坏得无法接受. 这类的统计方法或准则的研究属于**稳健统计**(robust statistics 又译作"鲁棒统计")的范畴.

非参数统计与稳健统计之间有紧密的联系：一些常用的非参数统计方法由于适用于相当大一类的分布模型，因而具有很好的稳健性. 本章前几节介绍常用的非参数统计方法，最后一节介绍稳健统计的基本概念.

14.2 顺序统计量及其应用

定义 14.2.1 设 $X_1,\cdots,X_n$ 为取自具连续分布 F 的总体的简单样本，将 $X_1,\cdots,X_n$ 重排为 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$(即 $X_{(i)}$ 为 $X_1,\cdots,X_n$ 中第 i 小的，$i=1,\cdots,n$)，当 $X_i=X_j,i<j$ 时，将 X_i 排在 X_j 前，则 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 称为由样本 $X_1,\cdots,X_n$ 生成的**顺序统计量**(order statistics).

定理 14.2.2 设总体分布函数为 $F(x)$，有密度 $f(x)$，$X_{(1)}$，$\cdots$，$X_{(n)}$为由简单样本生成的顺序统计量，则

(1) $(X_{(1)},\cdots,X_{(n)})$有联合密度

$$p(x_1,\cdots,x_n)=n!\ f(x_1)\cdots f(x_n),\quad 当\ x_1\leqslant x_2\leqslant\cdots\leqslant x_n.$$

(2) 对 $k\leqslant n$，及 $1\leqslant r_1<r_2<\cdots<r_k\leqslant n$，$(X_{(r_1)},\cdots,X_{(r_k)})$有密度

$$p_{r_1\cdots r_k}(x_1,\cdots,x_k)=n!\prod_{i=1}^{k}f(x_i)\prod_{i=1}^{k}\frac{[F(x_{i+1})-F(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!}.$$

$x_1\leqslant x_2\leqslant\cdots\leqslant x_k$，其中 $x_0\triangleq-\infty$，$x_{k+1}\triangleq+\infty$，$r_0\triangleq 0$，$r_{k+1}\triangleq n+1$.

(3) 对 $r\leqslant n$，$X_{(r)}$的密度为

$$p_r(x)=\frac{n!}{(r-1)!(n-r)!}F^{r-1}(x)[1-F(x)]^{n-r}f(x).$$

推论 14.2.3 在定理 14.2.2 的记号下有

(1) $X_{(1)}$的密度为

$$p_1(x)=n[1-F(x)]^{n-1}f(x),$$

(2) $X_{(n)}$的密度为

$$p_n(y)=nF^{n-1}(y)f(y),$$

(3) $(X_{(1)},X_{(n)})$的联合密度为

$$p(x,y)=n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y),\quad x\leqslant y.$$

推论 14.2.4 设 $F(x)$为连续总体分布函数，$X_{(1)},\cdots,X_{(n)}$为顺序统计量，记

$$Y_1=F(X_{(1)}),$$
$$Y_i=F(X_{(i)})-F(X_{(i-1)}),i=2,\cdots,n,$$
$$Y_{n+1}=1-F(X_{(n)}),$$

则$(Y_1,\cdots,Y_{n+1})\sim D_{n+1}(1,\cdots,1)$——狄利克雷分布(见节 6.6.3).

顺序统计量在非参数估计中有重要应用.

定义 14.2.5 设 $X_{(1)},\cdots,X_{(n)}$为顺序统计量，**经验分布函数**(empirical distribution function)为

$$F_n(x)=\begin{cases}0, & \text{当 } x \leqslant X_{(1)}, \\ \dfrac{i}{n}, & \text{当 } X_{(i)} < x \leqslant X_{(i+1)}, i=1,\cdots,n-1, \\ 1, & \text{当 } X_{(n)} < x.\end{cases}$$

对给定样本(因而给定了顺序统计量),经验分布函数 $F_n(x)$ 对于变量 x 是一个分布函数.另一方面,对固定的 x,$F_n(x)$作为样本的函数是一个统计量.

定理 14.2.6 格利文科(Гливенко)**定理** 设$\{X_n,n\geqslant 1\}$为独立同 F 分布随机变量列,$F_n(x)$为由 $X_1,\cdots,X_n$ 生成的经验分布函数,$n=1,2,\cdots$,则

$$P(\lim_{n\to\infty}[\sup_{-\infty<x<\infty}|F_n(x)-F(x)|]=0)=1.$$

此定理说明,用经验分布函数 $F_n(x)$直接去估计总体分布函数 $F(x)$,在整个实轴上有一致强相合性.由经验分布函数出发可派生出一系列有用的非参数估计.

一般,设待估参数为总体分布函数 F 的某一数字特征,记为 $T(F)$.在数学上,$T(F)$相当于定义在分布函数空间上的一个泛函.由于 F_n 是 F 的一致强相合估计,自然可以考虑用 F_n 的相应的数字特征 $T(F_n)$去估计 $T(F)$.作为一个重要的例子,矩估计便属于这类估计.根据矩估计的定义容易看出,样本矩即为经验分布的矩.因此矩估计方法的实质就是用经验分布的矩去估计总体分布相应的矩.从这个意义上讲,矩估计方法是一种非参数估计方法.但矩估计一般不具备常见的非参数统计方法的稳健性,而且从传统上,矩估计常用于参数估计的场合.

分位数估计是一种重要的基于经验分布函数的非参数估计方法.

定义 14.2.7 设 $X_{(1)},\cdots,X_{(n)}$为顺序统计量,对给定的 $p\in(0,1)$,用$[np]$记 np 的整数部分.

样本的下 p 分位数为

$$\xi_p(F_n) = \frac{1}{2}[X_{([np])} + X_{([np]+1)}],$$

特别当 $p=\frac{1}{2}$时，记 $\xi_{\frac{1}{2}}(F_n)=m(F_n)$，称为**样本中位数**.

类似地也可以定义样本的上 p 分位数. 为简便计，简称下 p **分位数为 p 分位数**.

定理 14.2.8 设对取自具有连续分布 F 的总体的简单样本 $X_1,\cdots,X_n$，记 $\xi_p(F_n)$为样本的 p 分位数，$\xi_p(F)$为总体的 p 分位数，$p\in(0,1)$，则

$$P(\lim_{n\to\infty}\xi_p(F_n) = \xi_p(F)) = 1.$$

此定理说明样本分位数为总体分位数的强相合估计.

定理 14.2.9 设总体分布 F 有连续密度 $f(x)$，给定 $0<p_1<\cdots<p_k<1$，用 ξ_i 记总体的 p_i 分位数，ξ_{ni}记样本的 p_i 分位数，$i=1,\cdots,k$. 记 $\xi=(\xi_1,\cdots,\xi_k)'$，$\xi_n=(\xi_{n1},\cdots,\xi_{nk})'$. 则

$$\lim_{n\to\infty}\sqrt{n}(\xi_n-\xi)\xlongequal{\text{d}}N_k(0,\Sigma),$$

其中，$\Sigma=(\sigma_{ij})_{k\times k}$，$\sigma_{ij}=\sigma_{ji}=p_i(1-p_j)/[f(\xi_i)f(\xi_j)]$.

推论 14.2.10 在定理 14.2.9 的假定下，记 $\xi_p(F)$和 $\xi_p(F_n)$分别为总体和样本的 p 分位数，则

$$\lim_{n\to\infty}\sqrt{n}(\xi_p(F_n)-\xi_p(F))\xlongequal{\text{d}}N(0,p(1-p)/f^2(\xi_p(F))).$$

特别，当 m 为总体的中位数，而 m_n 为样本中位数时，有

$$\lim_{n\to\infty}\sqrt{n}(m_n-m)\xlongequal{\text{d}}N(0,[4f^2(m)]^{-1}).$$

上述定理和推论说明了样本分位数为总体分位数的渐近正态估计.

样本分位数在估计总体分布的容忍区间和容忍限方面也有重要应用.

定义 14.2.11 设 F 为总体分布，$\hat{\theta}_1\leqslant\hat{\theta}_2$ 为两个统计量，若对两小正数 β,γ 有

(1) $P(F(\hat{\theta}_2)-F(\hat{\theta}_1)\geqslant 1-\gamma)\geqslant 1-\beta$,

则称$[\hat{\theta}_1,\hat{\theta}_2]$为 F 的一个**(β,γ)容忍区间**(tolerance interval).

(2) 若统计量$\hat{\theta}$满足

$$P(F(\hat{\theta})\geqslant 1-\gamma)\geqslant 1-\beta,$$

则称$\hat{\theta}$为 F 的**上(β,γ)容忍限**(tolerance limit).

(3) 若统计量$\hat{\theta}'$满足

$$P(F(\hat{\theta}')\leqslant \gamma)\geqslant 1-\beta,$$

则称$\hat{\theta}'$为 F 的**下(β,γ)容忍限**.

容忍区间和容忍限的估计有重要应用.例如在研究无线电元件的寿命时,人们希望知道 95%的元件的寿命在一个什么样的区间内.根据样本可以构造一个区间$[\hat{\theta}_1,\hat{\theta}_2]$.实际上人们不能期望$F(\hat{\theta}_1)-F(\hat{\theta}_2)\geqslant 95\%$总成立,而只能希望将$\hat{\theta}_1$和$\hat{\theta}_2$取得足够好,使得$P(F(\hat{\theta}_2)-F(\hat{\theta}_1)\geqslant 0.95)$足够大.用样本分位数就可以作到这点.

设 F 为连续总体分布,$X_1,\cdots,X_n$ 为简单样本.根据定理 6.5.3 知,$F(X_1),\cdots,F(X_n)$为独立$U(0,1)$分布的随机变量,因此$F(X_{(1)}),\cdots,F(X_{(n)})$为$U(0,1)$随机变量生成的顺序统计量,其分布与 F 无关.$F(X_{(k)})$有密度函数

$$f_k(x)=\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k},0\leqslant x\leqslant 1.$$

这是参数为$(k,n-k+1)$的 β 分布的密度.若 n 足够大,可取 k 尽可能大,使得

$$P(F(X_{(k)})\geqslant 1-\gamma)\geqslant 1-\beta,$$

则$X_{(k)}$为一个上(β,γ)容忍限.又可取 l 尽可能小使得

$$P(F(X_{(l)})\leqslant \gamma)\geqslant 1-\beta,$$

则$X_{(l)}$为一个下(β,γ)容忍限,这种计算需要查 β 分布表或编制专门的计算程序.

另外对 $l<k$,$(F(X_{(l)}),F(X_{(k)}))$有联合密度

$$f_{l,k}(x,y)=\frac{n!}{(l-1)!(k-l-1)!(n-k)!}$$
$$\times x^{l-1}(y-x)^{k-l-1}(1-y)^{n-k},0<x<y<1.$$
据此可通过计算以定出 l,k,使得
$$P(F(X_{(k)})-F(X_{(l)})\geqslant 1-\gamma)\geqslant 1-\beta.$$
于是 $(X_{(l)},X_{(k)})$ 即为一个 (β,γ) 容忍区间.

14.3 极值统计

在顺序统计量中,极大值统计量占有特殊的地位. 在生命统计和某些灾难性自然现象的研究中(例如研究洪水的规律),极大值统计量的分布及其特征是一个重要的课题. 本节介绍有关的重要概念及关于极限分布的一系列结果. 由于通过改变符号,很容易将极小值的问题转化为极大值的问题来研究,本节只讨论极大值的问题,并简称为**极值**(extreme).

定义 14.3.1 设 $F(x)$ 为一连续分布函数,$f(x)$ 为它的密度函数,相应的**强度函数**(intensity function)为
$$\mu(x)\triangleq\frac{f(x)}{1-F(x)}.$$

强度函数的意义为:设 $F(x)$ 为人的寿命分布,则 $\mu(x)\mathrm{d}x$ 相当于已知一个人活了 x 年而在 x 到 $x+\mathrm{d}x$ 时刻内将要死去的概率. 因此,强度函数 $\mu(x)$ 又称为**危险函数**(hazard function),而在可靠性统计中称为失效率(failure rate). $\mu(x)$ 为生命统计中的一个重要概念. 当 $\mu(x)$ 上升时,说明随着时光的流逝,人的死亡的概率增加;当 $\mu(x)$ 下降时,说明随着时光流逝,人的死亡概率下降. 许多类型的寿命(包括人或机器等),具有如下特点:在寿命的初期阶段,$\mu(x)$ 是下降的,而在寿命的晚期,$\mu(x)$ 上升.

由
$$\mu'(x)=\mu(x)\left(\frac{f'(x)}{f(x)}+\mu(x)\right),\tag{14.1}$$

可得如下的结论：

$$\mu'(x)\begin{cases}>0, & 当\ \mu(x)>-f'(x)/f(x),\\ =0, & 当\ \mu(x)=-f'(x)/f(x),\\ <0, & 当\ \mu(x)<-f'(x)/f(x).\end{cases} \quad (14.2)$$

定义 14.3.2 **临界商**(critical quotient)为

$$Q(x)\triangleq\frac{\mu(x)}{-f'(x)/f(x)}=\frac{-f^2(x)}{f'(x)[1-F(x)]}.$$

临界商 $Q(x)$是反映寿命分布的重要特征. 当 $Q(x)>1$ 时,随着生命的延续,继续存活的概率下降;当 $Q(x)<1$ 时,随着生命的延续,继续存活的概率上升.

定义 14.3.3 设 n 为正整数,$F(x)$为连续分布函数,相应的**特征极值**(characteristic extrem)u_n 满足

$$n(1-F(u_n))=1.$$

特征极值 u_n 的含义如下:设 F 为某一寿命分布函数,则 $1-F(x)$为某一个体寿命超过 x 的概率,而 $n(1-F(x))$为 n 个个体中寿命超过 x 的平均个数. u_n 为这样一个临界值:它使得平均每 n 个个体中寿命超过 u_n 的个数为 1.

定义 14.3.4 设 u_n 为分布函数 $F(x)$的特征极值,$\mu(x)$为强度函数,

$$\alpha_n\triangleq\mu(u_n)=\frac{f(u_n)}{1-F(u_u)}=nf(u_n),$$

称 α_n 为**极值强度**(extrem strength).

例 14.3.5 设 $F(x)=1-e^{-\lambda x}, x>0, \lambda>0$ 为参数. 则 $F(x)$为指数分布,易于计算:

$$\mu(x)\equiv\lambda,\qquad Q(x)\equiv 1.$$

$$u_n=\frac{1}{\lambda}\log n,\qquad \alpha_n=\lambda.$$

特征极值 u_n 和极值强度 λ_n 在讨论极值分布中是两个重要的

参数.

定义 14.3.6 设 $F(x)$ 为右无界的分布函数,即 $F(x)<1$, $\forall x$;其密度为 $f(x)$. 若 $F(x)$ 满足

$$\lim_{x\to\infty} Q(x) = 1,$$

则称 $F(x)$ 为**指数型分布**(distribution of exponential type).

常见的连续分布如指数分布,正态分布 Γ(Gamma)分布,对数正态分布及韦布尔(Weibull)分布等,都是指数型分布.

定理 14.3.7 设 $F(x)$ 为指数型分布,u_n 和 α_n 分别为特征极值和极值强度. 设 $X_1,\cdots,X_n$ 为取自 F 总体的独立同分布样本,$X_{(n)}$ 为最大值统计量(即 $X_{(n)}=\max(X_1,\cdots,X_n)$),则 $\alpha_n(X_{(n)}-u_n)$ 有极限分布函数

$$G(y) = \exp\{-e^{-y}\}, \ -\infty < y < \infty.$$

除定理 14.3.7 中给出的分布之外,任一分布 F 的最大值的极限分布还有两种可能性.

定理 14.3.8 设 $X_{(n)}$ 为独立同分布样本 $X_1,\cdots,X_n$ 的最大值统计量,若存在常数 d_n 及正常数 c_n 使得 $c_nX_{(n)}+d_n$ 依分布收敛于分布函数 $G(x)$,则 $G(x)$ 必与以下三个分布函数之一同型.

Ⅰ $G_1(x)=\exp(-e^{-x}), -\infty<x<\infty$;

Ⅱ $G_2(x)=\exp\left(-\dfrac{1}{x^k}\right)$, $x>0, k>0$ 为参数;

Ⅲ $G_3(x)=\exp\{-(-x)^k\}$, $x<0, k>0$ 为参数.

$G_1(x)$,$G_2(x)$ 和 $G_3(x)$ 分别称为第Ⅰ,Ⅱ,Ⅲ型极值分布.

极大值统计量的极值分布类型是由总体分布 $F(x)$ 的性质来决定的.

定义 14.3.9 设 $F(x)$ 为总体分布,$X_1,\cdots,X_n$ 为独立同分布样本,若极大值统计量 $X_{(n)}$ 的极限分布属于 i 型极值分布($i=$Ⅰ,Ⅱ,Ⅲ),则称 $F(x)$ 属于 i 型分布的**吸引场**(domain of attraction).

下面分别给出总体分布属于Ⅰ,Ⅱ,Ⅲ型极值分布吸引场的充

分必要条件.

定理 14.3.10 连续分布函数 $F(x)$ 属于 I 型极值分布吸引场的充分必要条件为

$$\lim_{n\to\infty} n\left[1 - F\left(u_n + \frac{x}{\lambda_n}\right)\right] = e^{-x}$$

其中 $F(u_n) = 1 - \frac{1}{n}, F\left(u_n + \frac{1}{\lambda_n}\right) = 1 - \frac{1}{ne}$.

进而若 $F(x)$ 有连续密度 $f(x)$，则有

$$\lim_{n\to\infty} \frac{\lambda_n}{nf(u_n)} = 1.$$

定理 14.3.11 连续分布 $F(x)$ 属于 II 型极值分布的吸引场的充分必要条件为

(1) $F(x) < 1$，一切 x；

(2) 存在 $k > 0$，使对任意 $c > 0$ 有

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - F(cx)} = c^k.$$

定理 14.3.12 连续分布 $F(x)$ 属于 III 型极值分布的吸引场的充分必要条件为

(1) 存在有限的 a，使得 $F(a) = 1$；

(2) 存在 $k > 0$，使对任何 c 有

$$\lim_{x\to 0} \frac{1 - F(cx + a)}{1 - F(x + a)} = c^k.$$

在实际问题中，通常假定总体分布是属于某个极值分布的吸引场. 这种假定是基于如下考虑：虽说由问题的非参数性质，我们不能断言总体分布有何种形式，但仍可以假定总体是(比如说)属于指数形的. 因此假定极值分布具第 I 型. 对独立同分布样本 $X_1, \cdots, X_n$，我们可将其分成 m 组，每组的个数相等且足够多. 用 $Y_1, \cdots, Y_m$ 分别记这 m 组中每组的最大值. 因此 $Y_1, \cdots, Y_m$ 可以近似地看成是从分布函数

$$G(y) = \exp\{-e^{-\alpha(y-u)}\}, \quad -\infty < y < \infty \tag{14.3}$$
中抽出的独立同分布样本，其中 α 和 u 为待估的参数.

根据上述理由，我们可利用极大似然估计的方法由 $Y_1,\cdots,Y_m$ 去估计 α 和 u.

下面介绍一个简单的方法.

方法 14.3.13 利用样本分位数估计参数

由
$$G(u) = \exp\{-e^0\} = e^{-1} = 0.3679 = p_1,$$
及
$$G\left(u + \frac{1}{\alpha}\right) = \exp\{-e^{-1}\} = 0.6922 = p_2.$$
可用 $Y_1,\cdots,Y_m$ 的样本 p_1 分位数 ξ_{p_1} 去估计 u，而用 p_2 分位数 ξ_{p_2} 去估计 $u+1/\alpha$. 并由此解得
$$\hat{u} = \xi_{p_1}, \hat{\alpha} = \frac{1}{\xi_{p_2} - \xi_{p_1}}.$$
此方法简便易行，但精度不如极大似然估计.

14.4 秩检验

在非参数估计问题中，顺序统计量及其函数有着重要的作用. 但在非参数假设检验问题中，由于涉及到计算概率的问题，而顺序统计量的分布是与总体分布有关的，因此直接由顺序统计量去构造检验函数的效果不太好. 本节介绍一类在非参数检验问题中有广泛应用的统计量——**秩统计量**(rank statistics). 秩统计量的一个重要特性是**分布无关性**(distribution-freeness).

定义 14.4.1 设 $X_1,\cdots,X_n$ 为一组样本(不必是取自同一总体的). 将 $X_1,\cdots,X_n$ 从小到大排成一列，用 R_i 记 X_i 在上述排列中的位置序号，$i=1,\cdots,n$. 称 $R_1,R_2,\cdots,R_n$ 为由样本 $X_1,\cdots,X_n$ 产

生的**秩统计量**(rank statistics).

例 14.4.2 例如有下列一组样本：

x_1	x_2	x_3	x_4	x_5
1.2	0.8	−3.1	2.0	1.2

则由此样本产生的秩统计量为：$R_1=3, R_2=2, R_3=1, R_4=5, R_5=4$.

注意在上述例中，$x_1=x_5$，这时我们按照自然顺序将 x_1 排在 x_5 之前.

显然，若样本 $X_1,\cdots,X_n$ 是取自连续分布总体的独立同分布样本，则秩统计量 $R_1,\cdots,R_n$ 的分布是对称等概的. 即对 $1,\cdots,n$ 的任一排列 $i_1,\cdots,i_n$，有

$$P(R_1=i_1, R_2=i_2, \cdots, R_n=i_n)=\frac{1}{n!} \tag{14.4}$$

这时，$R_1,\cdots,R_n$ 的分布与总体分布无关.

14.4.1 斯皮尔曼秩相关检验

设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 为取自某个二元总体的独立样本，要检验 X 变量与 Y 变量是否相关. 以"X 与 Y 相互独立(不相关)"为原假设，对立假设可能有不同的选择. 例如假设"X 与 Y 正相关"或"X 与 Y 负相关"，或更一般地，假设"X 与 Y 相关"，在正态总体假定下，最自然的构造检验的办法是用样本相关系数

$$\begin{aligned}\hat{\rho}_{xy}&=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})\Bigg/\sqrt{\frac{1}{n}\sum_{1}^{n}(X_i-\overline{X})^2\frac{1}{n}\sum_{1}^{n}(Y_i-\overline{Y})^2}\\&=\left[\frac{1}{n}\sum_{i}(X_iY_i-\overline{X}\,\overline{Y})\right]\Bigg/\sqrt{\frac{1}{n}\sum_{1}^{n}(X_i-\overline{X})^2\frac{1}{n}\sum_{1}^{n}(Y_i-\overline{Y})^2}\end{aligned} \tag{14.5}$$

而在非参数的相关性检验中，斯皮尔曼(Spearman)首先提出了**秩相关系数**(rank correlation coefficient)的概念．设$(r_1,\cdots,r_n)$为由$X_1,\cdots,X_n$产生的秩统计量，$R_1,\cdots,R_n$为由$Y_1,\cdots,Y_n$产生的秩统计量

$$\bar{r} \triangleq \frac{1}{n}\sum_1^n r_i = \frac{n+1}{2} = \bar{R} \triangleq \frac{1}{n}\sum_1^n R_i,$$

$$\frac{1}{n}\sum_1^n (r_i - \bar{r})^2 = \frac{n^2-1}{12} = \frac{1}{n}\sum_1^n (R_i - \bar{R})^2. \quad (14.6)$$

定义 14.4.3 斯皮尔曼**秩相关系数**(Spearman rank correlation coefficient)为

$$r_s = \left[\frac{1}{n}\sum_1^n r_i R_i - \left(\frac{n+1}{2}\right)^2\right] \Big/ \left(\frac{n^2-1}{12}\right).$$

当X与Y相互独立时，$(r_1,\cdots,r_n)$与$(R_1,\cdots,R_n)$是相互独立的，此时$E(r_s)=0$．当X与Y正相关时，r_s倾向于取正值；而当X与Y负相关时，r_s倾向于取负值．这就提示了我们利用r_s来构造检验的方法．问题是计算r_s的分布．

定理 14.4.4 当n较大时，$\sqrt{n-1}\,r_s$的近似分布为$N(0,1)$——标准正态分布．

因此，对假设检验问题

H_0: X与Y不相关$\leftrightarrow H_1$: X与Y正相关． (14.7)

基于斯皮尔曼秩相关系数的检验方法为，选k，使得

$$P_{H_0}(r_s > k) \leqslant \alpha, \quad (14.8)$$

其中α为指定的检验水平，P_{H_0}是表示在原假设H_0真的条件下求概率，则

当$r_s > k$时否定原假设；

当$r_s \leqslant k$时肯定原假设． (14.9)

在上述检验方法中，k的确定可利用定理14.3.5，即取k满足

$$\Phi(\sqrt{n-1}\,k) = 1-\alpha, \quad (14.10)$$

其中 Φ 为标准正态分布函数，Φ 的值可通过查正态分布函数表来求得.

类似可构造负相关检验及一般的相关性假设的双边检验.

斯皮尔曼秩相关系数的另一表达形式为

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_1^n (R_i - r_i)^2, \tag{14.11}$$

当 X 变量与 Y 变量相互独立时，其均值和方差分别为

$$E(r_s) = 0, \quad v_{ar}(r_s) = (n-1)\left(\frac{n-1}{12}\right)^2.$$

14.4.2 随机性检验

设 $X_1,\cdots,X_n$ 为一组按顺序抽得的样本，这顺序可能是时间或空间上的顺序或根据某种设计好的抽样方法所规定的顺序. 所谓**随机性**(randomness)，是指样本值的大小与抽样的顺序无关. 而与此对立的假设是认为样本取值大小随抽取的顺序而表现出一定的倾向性. 譬如说，样本值随抽样顺序而逐渐有增大的趋势. 用数学方法来表示，设 $F_i(x)$ 为样本 X_i 的分布函数，$i=1,\cdots,n$，则上述两种对立的假设可表示为：

$$H_0: F_1(x)=\cdots=F_n(x) \leftrightarrow H_1: F_1(x)>\cdots>F_n(x),$$

对任意的 x. (14.12)

为构造假设检验问题式(14.12)的检验，考虑样本 $X_1,\cdots,X_n$ 所产生的秩统计量 $R_1,\cdots,R_n$. 当 $i<j$，而 $R_i>R_j$ 时称为有一个逆序. 记

$$Q \triangleq R_1,\cdots,R_n \text{ 中所有逆序的个数}. \tag{14.13}$$

例如设$(R_1,\cdots,R_5)=(3,2,5,4,1)$，则其中有 6 个逆序：3—2，3—1，2—1，5—4，5—1，4—1. 显然，Q 最小值取零(此时$(R_1,\cdots,R_n)=(1,\cdots,n)$)；最大值取$\frac{1}{2}n(n-1)$(此时$(R_1,\cdots,R_n)=(n,\cdots,1)$).

定义

$$t \triangleq 1 - \frac{4Q}{n(n-1)}. \tag{14.14}$$

显然$-1 \leqslant t < 1$，且当式(14.12)中H_0成立时，t关于原点有对称分布；而当H_1成立时，t的分布偏向于1. 关于t的近似分布有以下结论.

定理 14.4.5 设$X_1, \cdots, X_n$为独立同分布样本，则当n较大时，$t\Big/\sqrt{\dfrac{2(2n+5)}{9n(n-1)}}$的分布近似于标准正态分布$N(0,1)$.

对$n \geqslant 5$，上述近似分布有较好的精确度，而对$n \leqslant 4$可直接计算t的分布.

例 14.4.6 计算$n=4$时t的分布.

$$P(Q=0)=P(t=1)=\frac{1}{24},$$

$$P(Q=1)=P\left(t=\frac{2}{3}\right)=\frac{3}{24}=\frac{1}{8},$$

$$P(Q=2)=P\left(t=\frac{1}{3}\right)=\frac{5}{24},$$

$$P(Q=3)=P(t=0)=\frac{6}{24}=\frac{1}{4},$$

$$P(Q=4)=P\left(t=-\frac{1}{3}\right)=\frac{5}{24},$$

$$P(Q=5)=P\left(t=-\frac{2}{3}\right)=\frac{3}{24}=\frac{1}{8},$$

$$P(Q=6)=P(t=-1)=\frac{1}{24}.$$

对假设检验问题式(14.12)，基于统计量t的水平α的检验可构造如下：取t_0满足

$$P_{H_0}(t > t_0) = \alpha,$$

其中，P_{H_0}表示在原假设H_0成立的条件下求概率，此概率值可根

据定理 14.4.5 所提供的近似分布计算,则相应的检验为:

当 $t>t_0$,则否定 H_0,肯定 H_1;

当 $t\leqslant t_0$,则否定 H_1,肯定 H_0.

14.4.3 两样本齐次性检验

考虑在两台同类仪器上做同一种试验,分别得到两组数据,现在怀疑其中一台仪器有偏差,因此测得的数据比另一台仪器上测得的数据有偏大的倾向.用数学的语言表达如下:设有两组样本:$X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$,用 $F(x)$和 $G(x)$分别表示 X 和 Y 的分布函数.考虑下述的假设检验问题:

$$H_0:F(x)=G(x),\text{一切 } x,\leftrightarrow H_1:F(x)>G(x),\text{一切 } x. \quad (14.15)$$

式(14.15)是一个**齐次性**(homogeneity)假设检验问题.利用秩统计量构造检验的方法如下:将样本 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$ 合并成一个大的样本,用 $R_1,\cdots,R_m$ 记 $Y_1,\cdots,Y_m$ 在整个合并样本当中的秩.R_i 的取值范围为 1 到 $n+m$,$i=1,\cdots,m$.当 H_0 成立时,$X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$ 可看成是从同一总体中抽取的容量为 $n+m$ 的独立样本,这时 $R_1,\cdots,R_m$ 应该比较均匀地分布在 1 到 $n+m$ 之间;而当 H_1 成立时,Y 样本的取值比较偏大,因此 $R_1,\cdots,R_m$ 的分布偏向上述区间的右端.根据上述思想,威尔柯克森(Wilcoxon)提出了如下的统计量.

定义 14.4.7

$$w \triangleq \sum_{i=1}^{m} R_i$$

称为威尔柯克森**秩和统计量**(Wilcoxon rank sum statistics).

当 H_0 成立即 $X_1,\cdots,X_n,Y_1,\cdots,Y_m$ 独立同分布时,有

$$E(w) = \frac{1}{2}m(n+m+1),$$

$$\mathrm{v_{ar}}(w) = \frac{1}{12}mn(n+m+1).$$

定理 14.4.8 当(14.15)中 H_0 成立时,设 $n+m\to\infty, n/m\to\lambda, 0<\lambda<\infty$,则

$$\left(w-\frac{1}{2}m(n-m+1)\right)\Big/\sqrt{\frac{1}{12}nm(n+m+1)}$$

的渐近分布为标准正态分布 $N(0,1)$.

假设检验问题(14.15)的基于 w 的检验可构造如下:取 w_0,使得

$$P_{H_0}(w>w_0)\leqslant\alpha,$$

其中 α 为给定的检验水平,上述概率可利用定理 14.4.8 近似求出,于是

当 $w>w_0$ 时,否定 H_0,肯定 H_1;

当 $w\leqslant w_0$ 时,肯定 H_0,否定 H_1.

14.5 核方法与近邻方法

本节介绍两种在估计总体的局部特征方面很有用的非参数方法及其在非参数密度估计、非参数回归和非参数判别中的应用.

14.5.1 密度估计

设样本 $X_1,\cdots,X_n$ 是从具有未知密度 $f(x)$ 的一维总体中抽出的简单样本,要依据这些样本去估计 $f(x)$.

最古老而又简便的密度估计方法就是**直方图**(histogram),其方法可描述如下.

方法 14.5.1 直方图密度估计 将整个实轴分成 m 个小区间 $[a_i,a_{i+1}), i=1,\cdots,m$(通常在 $[a_1,a_{m+1}]$ 内包含了全部或绝大部分样本点),设在区间 $[a_i,a_{i+1})$ 内有 n_i 个样本($X_1,\cdots,X_n$ 中有 n_i 个落在区间 $[a_i,a_{i+1})$ 内).根据频率逼近概率的思想,可用 n_i/n 去估计总体分布在区间 $[a_i,a_{i+1})$ 上的概率,而在区间 $[a_i,a_{i+1})$ 上的概

率密度估计可用

$$\hat{f}_n(x) = \frac{n_i}{n\Delta_i}, \text{当 } x \in [a_i, a_{i+1}), \tag{14.16}$$

去估计，其中 $\Delta_i = a_{i+1} - a_i$，为区间长度.

直方图密度估计具有直观易懂的特点，是对总体分布进行初步分析的一种简便有效的工具. 在作直方图时，区间划分的方式需要根据实际情形和经验来决定. 一般，这 m 个区间应包含全部或绝大部分(比如说 99%)的样本. 另外 m 的大小应根据样本容量 n 的大小而定. m 过小，则每个区间较大，因此造成在区间中心部分估计效果较好，而在区间端点部分估计效果较差；m 过大，则落在每个区间内样本点较少，使得估计偏差较大. 作为供读者参考的一种明确说法，建议取 $m = \sqrt{n}$ 的整数部分. 图 14.1 给出了直方图的一个示意图.

为克服直方图在区间中心部分估计效果较好，而在区间端点附近效果较差的缺点. 一种改进的方法是：对每个 x，以 x 为中心作区间 $[x-h, x+h]$，其中 h 为事先选定的区间半长度. 用 n_x 记样本 $X_1, \cdots, X_n$ 落入此区间的个数，用

$$\hat{f}_n(x) = \frac{n_x}{2hn}, \tag{14.17}$$

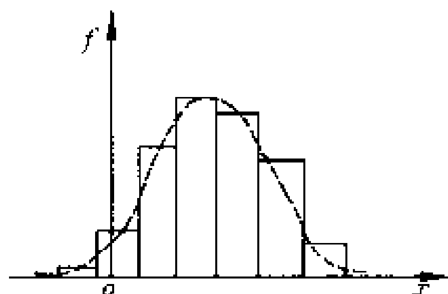


图 14.1　直方图
(实线的并列矩形为直方图，虚线为假定的总体密度曲线).

去估计密度 $f(x)$. 由于对每一个 x 值，都是以它为中心去构造区间，显然这种密度估计的效果优于直方图估计，且具有较好的连续性. (14.17)所给出的估计可用另一种形式表示. 设

$$K(x) = \begin{cases} \frac{1}{2}, & \text{当 } -1 \leqslant x \leqslant 1, \\ 0, & \text{其它}, \end{cases} \tag{14.18}$$

则 $K(x)$为$[-1,1]$区间上均匀分布的密度函数,而(14.17)式可改写为

$$\hat{f}_n(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right). \tag{14.19}$$

在上式中,将 $K(x)$换成任意一个密度函数就得到密度的核估计.

定义 14.5.2 设 $K(x)$为一概率密度函数,$h>0$ 为给定常数,$f(x)$为总体密度,则形如(14.19)式的$\hat{f}_n(x)$称为 $f(x)$的一个**核估计**(kernal estimate),其中函数 $K(x)$称为**核**(kernal),常数 h 称为**窗宽**(window-width).

核密度估计的实质是对样本点施以不同的权,用加权来代替通常的计数.在实际应用中,核函数 K 常用在原点有单峰的对称密度函数.除了(14.18)式中给出的均匀核外,还可考虑用

$$K(x)=\begin{cases}\frac{1}{2}\cos x, & \text{当 } |x|\leqslant\frac{\pi}{2},\\ 0, & \text{当 } |x|>\frac{\pi}{2}.\end{cases} \tag{14.20}$$

$$K(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad -\infty<x<+\infty. \tag{14.21}$$

窗宽 h 的选择与 n 有关,太大或太小都不好,但没有什么明确的优选准则.理论上说,当 $n\to\infty$时,要求 $h\to0$,但 $nh\to\infty$.对固定的 n,h 的选择可根据具体情况和经验试几个值,从中选一个较为满意的.作为一个建议,可先试 $h=1/\sqrt{n}$.

核密度估计可推广到多维总体的情形.

定义 14.5.3 设 $X_1,\cdots,X_n$ 为取自 d 维连续总体的简单样本,$K(x)$为 d 维密度函数,$h>0$ 为常数,则总体密度 $f(x)$的一个核估计可表为

$$\hat{f}_n(x)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right).$$

在定义(14.17)中，区间$[x-h,x+h]$的半长度 h 为事先给定的，而落入此区间中的样本数 n_x 是随机的. 按相反的思路，可考虑固定落入区间中的样本数，而令区间半长度为随机的，这样就得到密度的近邻估计.

定义 14.5.4 设 $X_1,\cdots,X_n$ 为取自一维连续总体的简单样本，对指定的实数 x 及正整数 k，设 X_{R_k}为距 x 第 k 个最近的样本点. 令

$$\hat{f}_n(x)=\frac{k}{2n|x-X_{R_k}|},$$

则称$\hat{f}_n(x)$为总体密度 $f(x)$的**近邻密度估计**(nearest neighbors density estimate).

近邻密度估计与(14.17)给出的均匀核密度估计相比，前者在各处估计的精度有较好的一致性，即在各处的估计精度相差不多；而后者在样本密集的地方估计精度高，在样本稀疏的地方则估计精度差.

在近邻密度估计中，入选的近邻样本点个数 k 的选择是一个关键. 理论上要求当 $n\to\infty$时，$k\to\infty$，而 $k/n\to 0$. 在实用上，作为一个供参考的建议，可考虑取 $k=\sqrt{n}$ 的整数部分.

近邻估计与核估计相结合，可得到一种推广形式为

$$\hat{f}_n(x)=\frac{1}{n|x-X_{R_k}|}\sum_{i=1}^{n}K\left(\frac{x-X_i}{|x-X_{R_k}|}\right),\qquad(14.22)$$

其中 K 为核，X_{R_k}为 x 的第 k 个近邻样本点.

关于核密度估计和近邻密度估计的大样本性质，有以下一些基本结论.

定理 14.5.5 对总体密度 $f(x)$的核估计 $\hat{f}_n(x)$(定义 14.5.2)，记窗宽为 h_n(因与 n 有关)，若

(1) 核函数 K 在$(-\infty,+\infty)$上有界，

(2) $f(x)$在$(-\infty,+\infty)$上有界，

(3) $\lim\limits_{n\to\infty} h_n = 0$,

则在 f 的连续点 x 有

$$\lim_{n\to\infty} E(\hat{f}_n(x)) = f(x).$$

在上述结论中,$E(\hat{f}_n(x))$是将$\hat{f}_n(x)$看成是样本 $X_1,\cdots,X_n$ 的函数,因而作为随机变量求期望的.

定理 14.5.6 在定理 14.5.5 的记号及条件下,若进一步有 $\lim\limits_{n\to\infty}(nh_n)=\infty$,则对 f 的连续点 x 有

$$\lim_{n\to\infty} E[\hat{f}_n(x) - f(x)]^2 = 0,$$

因而有

$$\hat{f}_n(x) \xrightarrow{\text{P}} f(x), \quad n\to\infty$$

定理 14.5.7 对总体密度 $f(x)$的近邻估计$\hat{f}_n(x)$(定义 14.5.4),记入选近邻个数为 k_n(因与 n 有关),则当

(1) $\lim\limits_{n\to\infty} k_n = \infty$,

(2) $\lim\limits_{n\to\infty}(k_n/n) = 0$

时,有

$$\hat{f}_n(x) \xrightarrow{\text{P}} f(x), \quad n\to\infty.$$

14.5.2 非参数回归

设 X 和 Y 分别是 d 维和一维随机变量. 若 $E|Y|<\infty$,则 $m(x)=E(Y|X=x)$存在,称为 Y 对 X 的回归函数. 回归分析的基本问题,就是用(X,Y)的简单样本$(X_1,Y_1),\cdots,(X_n,Y_n)$去估计回归函数 $m(x)$.

在线性回归模型中,假定 $m(x)$是 x 的线性函数,用最小二乘法对回归系数进行估计(见第 15 章). 但是,经验和理论都证明,当 $m(x)$不是 x 的线性函数时,基于最小二乘的回归效果不好. 非参数回归就是在对 $m(x)$的形式不作任何假定的前提下研究估计

$m(x)$的方法.

由样本均值估计总体均值的思想出发,假设样本(X_1,Y_1),$\cdots$,(X_n,Y_n)中有相当多的X_i恰好等于指定的x,记为X_{i_1},X_{i_2},$\cdots$,X_{i_k},这时自然可取相应的Y样本$Y_{i_1},\cdots,Y_{i_k}$,用它们的算术平均$\frac{1}{k}\sum_{j=1}^{k}Y_{i_j}$去估计$m(x)$.可是在实际问题中一般不会有很多$X$的样本值恰好等于指定的$x$(甚至一个也没有).但由此可启发我们在思路上产生一个飞跃,即用$Y_1,\cdots,Y_n$的一个加权和

$$\hat{m}_n(x)=\sum_{i=1}^{n}W_iY_i \tag{14.23}$$

去估计$m(x)$,其中$W_i\geqslant 0,i=1,\cdots,n$,$\sum_{i=1}^{n}W_i=1$.$\{W_i\}$称为一组"权"(weight).在选择权$\{W_i\}$时要考虑到:当$X_i=x$或$X_i$与$x$非常接近时,相应的权$W_i$应较大;反之则$W_i$应较小甚至为0.

定义 14.5.8 由(14.23)给出的$\hat{m}_n(x)$称为$m(x)$的权函数估计.

根据上述关于权的选取原则的思想,知道权$\{W_i,i=1,\cdots,n\}$是x和样本$X_1,\cdots,X_n$的一组函数,常用的权函数有核函数权和近邻权.

定义 14.5.9 选定R^d上的非负函数K及正常数h,令

$$W_i=K\left(\frac{x-X_i}{h}\right)\Bigg/\sum_{j=1}^{n}K\left(\frac{x-X_j}{h}\right),\quad i=1,\cdots,n,$$

则$\{W_i,i=1,\cdots,n\}$称为一组**核函数权**.

在上述定义中,核K不必是密度函数,但常用的核为在原点有单峰的对称函数.上节推荐的三种核函数(14.19),(14.20),(14.21)可供考虑,窗宽h的选择可参见14.5.1中有关叙述.

关于近邻权,事先应在R^d中引进一个适当的距离函数$\|\cdot\|$,即对u和$v\in R^d$,$\|u-v\|$表示u和v两点的距离.对指

定的 x，将样本 $X_1,\cdots,X_n$ 按到 x 的距离重排为 $X_{R_1},\cdots,X_{R_n}$，使得

$$\| x-X_{R_1} \| \leqslant \| x-X_{R_2} \| \leqslant \cdots \leqslant \| x-X_{R_n} \| . \quad (14.24)$$

定义 14.5.10 设 $c_1,\cdots,c_n$ 为一组非负实数，满足 $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n$ 且 $\sum_{i=1}^{n} c_i = 1$. 在定义 14.5.9 所规定的意义下，令

$$W_{R_i} = c_i, i = 1,\cdots,n.$$

则 $\{W_i, i=1,\cdots,n\}$ 称为一组**近邻权**(nearest neiburgh weights)

选取近邻权的基本思路是：与 x 相近的 X_i 所对应的 Y_i 赋予较大的权. 一种常用的近邻权是：选定一个正整数 k，令 $c_1=c_2=\cdots=c_k=\frac{1}{k}$，$c_{k+1}=\cdots=c_n=0$. k 的选取准则见节 14.5.1 中关于近邻密度估计中的有关叙述.

由于 X 是 d 维的，所以距离函数的选择有一定的任意性. 通常可考虑一种加权的闵可夫斯基(Minkowski)距离，即对 R^d 中的两个点 $u=(u_1,\cdots,u_d)$ 和 $v=(v_1,\cdots,v_d)$，令

$$\| u - v \|_r = \left[\sum_{l=1}^{d} a_l |u_l - v_l|^r \right]^{\frac{1}{r}}, \quad (14.25)$$

其中 $r \geqslant 1$ 为闵可夫斯基常数，$a_l > 0, l=1,\cdots,d$，$\sum_{l=1}^{d} a_l = 1$. $\{a_l\}$ 为事先给定的一组权，赋予各个分量以不同的重要性或将各分量的单位统一化.

另外还可考虑使用最大分量模距离.

$$\| u - v \|_\infty = \max_{1 \leqslant l \leqslant d} a_l |u_l - v_l|, \quad (14.26)$$

其中 a_l 的含义如上.

除了用权函数估计直接去拟合回归函数 $m(x)$ 之外，还可将权函数方法与最小二乘法结合起来使用. 具体作法是这样的：首先由样本 $(Y_1,X_1),\cdots,(Y_n,X_n)$ 去拟合一个线性模型 $y=\beta_0+\beta_1' x$，用最小二乘法得到线性回归系数 β_0 和 β_1(d 维向量)的估计 $\hat{\beta}_0$ 和

$\hat{\beta}_1$. 然后令

$$Z_i = Y_i - \hat{\beta}_0 - \hat{\beta}'_1 X_i, i = 1, \cdots, n, \tag{14.27}$$

再对$(Z_1, X_1), \cdots, (Z_n, X_n)$用权函数方法，得到一个非参数回归估计 $\widetilde{m}_n(x)$(相当于在式(14.25)中用 Z_i 去代替 Y_i). $\widetilde{m}_n(x)$和形式上看成是变量 Z 关于变量 X 的回归函数的估计. 这时整个回归函数(Y 关于 X 的)估计为

$$\hat{m}_n(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \widetilde{m}_n(x). \tag{14.28}$$

这样做的好处是：既给出了回归函数的线性部分$(\hat{\beta}_0 + \hat{\beta}_1 x)$，同时又对其非线性部分(如果确实存在的话)给出了一个合理的估计$(\widetilde{m}_n(x))$.

非参数回归也如同非参数密度估计一样，需要在实践中反复使用才能积累可贵的经验.

14.5.3 非参数判别

关于判别分析的一般概念，在第18章第5节中已有较为详细的介绍. 假定总体分为 m 个子体 $\Pi_1, \cdots, \Pi_m$，我们手头有一批样本$(Y_1, X_1), \cdots, (Y_n, X_n)$，通常称为**训练样本**(training sample)其中 X_i 为 d 维的，表示第 i 个样品的有关指标值，Y_i 取值于$\{1, \cdots, m\}$，表示第 i 个样品是取自子体的编号($Y_i = j$ 表示第 i 个样品取自子体 Π_j). 现在来了一个新的样品，其指标值为 X，要判别其所属的子体. 用 Y 记其所属子体的编号，问题转化为估计 Y. 下面介绍用核方法和近邻方法进行非参数判别的一般程序.

首先介绍用核方法进行判别. 设 π_j 为子体 Π_j 在总体中所占的比例，$f_j(x)$为子体 Π_j 的密度，$j = 1, \cdots, m$. 根据定理 18.5.3 知，贝叶斯判别法则是：

$$\text{当 } \pi_l f_l(x) = \max_{1 \leqslant j \leqslant m} \pi_j f_j(x) \text{ 时，判定 } Y = l. \tag{14.29}$$

在非参数判别中，通常都假定 $f_j(x), j = 1, \cdots, m$ 完全未知，有时也假定 $\pi_1, \cdots, \pi_m$ 未知. 因此，一个自然的想法是：直接估计

$f_j(x), j=1,\cdots,m$ 和 $\pi_j, j=1,\cdots,m$. 然后将得到的估计值代入式(14.28),便可进行判别. 具体作法是:将训练样本按 Y_i 的取值分组,使得 Y_i 取相同值的样本归于一组.(通常要求训练样本在 m 个子体内都有代表. 否则若某个子体内没有训练样本,则无法判别新样本是否属于该子体.)为叙述简便计,不妨设 $m=2, Y_1=Y_2=\cdots=Y_{n_1}=1, Y_{n_1+1}=\cdots=Y_n=2$. 于是将 X 样本分成两组:$\{X_1,\cdots,X_{n_1}\}$, $\{X_{n+1},\cdots,X_n\}$. 前一组取自 Π_1,后一组取自 Π_2. 用这两组训练样本分别对 $f_1(x)$ 和 $f_2(x)$ 按照定义 14.5.3 作核密度估计如下:

$$\hat{f}_1(x) = \frac{1}{n_1 h^d}\sum_{i=1}^{n_1} K\left(\frac{x-X_i}{h}\right),$$

$$\hat{f}_2(x) = \frac{1}{n_2 h^d}\sum_{i=n_1+1}^{n} K\left(\frac{x-X_i}{h}\right), \qquad (14.30)$$

其中 K 和 h 分别为选定的核和窗宽,$n_2=n-n_1$.

至于 π_1 和 π_2 的估计可用

$$\hat{\pi}_1 = n_1/n, \hat{\pi}_2 = n_2/n. \qquad (14.31)$$

使用这估计需要有一个前提:全部训练样本是从大的总体中随机抽取的;否则便没有意义.

将 $\hat{f}_1, \hat{f}_2$ 和 $\hat{\pi}_1, \hat{\pi}_2$ 代入(14.29)式,便可进行判别.

至于近邻方法用于非参数判别则有两条路径可循,其一是如同上述用核方法估计 $f_1(x)$ 和 $f_2(x)$ 一样,用近邻法去估计 $f_j(x)$,然后代入(14.29)进行判别. 另一条路径是直接法,其做法如下:在 R^d 中定义一个距离函数 $\|\cdot\|$(见上一小节中的有关叙述). 设新样本 $X=x$. 根据(14.24)式将样本重排为 $X_{R_1}, X_{R_2},\cdots,X_{R_n}$. 指定一个正整数 k,在对应着 $X_{R_1},\cdots,X_{R_k}$ 的 Y 样本值 $Y_{R_1},\cdots,Y_{R_k}$ 中,按照多数选举法的原则来判定 Y 值. 即当 j 为 $Y_{R_1},\cdots,Y_{R_k}$ 中出现的次数最多的值,则判定 $Y=j$. 若 $Y_{R_1},\cdots,Y_{R_k}$ 中出现最多的值不止一个,则判定为序号最小的. 这样做的一个前提是事先将子体 Π_1,

$\cdots,\Pi_m$ 按照一旦错判后造成的严重性从大到小进行排序.比如说,病人有病而错判为无病所造成的后果比无病而错判为有病的后果更为严重,因此代表病人有病的子体序号应在前,而代表无病的子体序号应在后.

以上非参数判别方法的优点是:对总(子)体的分布形式事先不作任何假定,因而有较广泛的适应性.特别是最后介绍的近邻判别法,它避开了对 π_i 的估计,在实际应用中更有其优点.

如同在密度估计和回归估计中的情形一样,在非参数判别中用核方法时核和窗宽的选择及在用近邻法时距离函数及近邻个数的选择都要根据具体情况和经验来确定,没有一成不变的规则.

14.6 稳健统计

统计学的基本方法是对数据建立一定的概率模型,就是说样本服从一定的概率分布,在此基础上根据问题的要求(估计某一参数,检验某一假设等)提出一定的统计方法(一个点估计量,一个检验等).对同一个统计问题的不同方法,通常要提出一个适当的择优准则(例如在点估计问题中要求方差最小,无偏估计等),在此准则下去选择"最优"的统计方法.

但在实际问题中常遇到的情形是:人为建立起来的数学统计模型与实际背景不可能完全一致,而是存在着一定的(尽管可能是微小的)偏差.例如,通常总假定样本是采自正态总体,而实际上总体分布可能是一种接近正态而非正态的分布.另外,在采集数据的过程中由于某种失误(例如仪器工作不正常,或记录数据时发生笔误等),常使样本中的若干个数据不真实.这些失真的数据常以离开数据群体而孤立出现,通常称为**异常值**(outlier).当上述情形出现时,某些在假定的模型下具有理论上优良性质的统计方法会变得效果很坏!例如在正态总体下,样本均值 $\overline{X}$ 是总体均值的最小

方差无偏估计，但当样本中混入一个很大的异常值时，$\overline{X}$ 的变化很 大，导致估计严重失真.

基于上述考虑，人们提出对统计方法要求有**稳健性**（robustness，又译为鲁棒性），其含义为：(1) 在实际背景与所假定的统计模型符合时，该统计方法具有良好的性能（但不必是在某种准则下最优）；(2) 当实际背景与假定模型有少许差异时，其性能所受到的影响也较小；(3) 当实际背景与假定模型严重不符时，其性能不致太坏. 满足上述要求的统计方法便称为具有**稳健性**.

在数学上描述稳健性，统计学家定义了一些概念，并在此基础上建立起一套理论. 本书对一些常用的方法作些介绍，本节的内容是关于稳健统计的若干片段.

首先介绍总体均值和总体方差的稳健估计. 熟知，样本均值和样本方差作为总体均值和方差的估计，即使在非参数估计的场合（即假定总体分布完全未知，只承认总体均值和方差均存在且有限），也是常用的估计方法，具有强相合性和渐近正态性，但是其前提必须是样本为独立同分布的. 当样本受到污染，即混入一定的异常值时，样本的独立同分布性便被破坏了，这时用样本均值和样本方差去估计总体均值和总体方差的效果就会变得很坏. 由于异常值通常表现得比正常的数据偏大（或偏小），因此造成均值的估计会偏大（或偏小）. 一个自然的想法是：将样本中特别大或特别小的数据（有一定理由怀疑它们是异常值）去掉，用剩下的数据进行算术平均来估计总体均值. 这种做法并不是什么特别新鲜的东西. 例如在体育比赛中，当由多个裁判员用记分的方式来决定名次时，为避免个别裁判不公正地给运动员打过高或过低的分，从而严重地影响运动员的名次，一种常见的做法是：去掉一个最高分及一个最低分，用剩下的中间分进行平均，这便是运动员的实得分数. 这就是所谓的“切尾均值”(trimmed mean).

定义 14.6.1 设 $X_1,\cdots,X_n$ 为样本，$X_{(1)},\cdots,X_{(n)}$ 为相应的顺

序统计量，给定正数 $\alpha<0.5$，**切尾均值**为

$$T_n=\frac{1}{n(1-2\alpha)}\sum_{i=[n\alpha]}^{[n(1-\alpha)+1]}X_{(i)}.$$

切尾均值的含义是：将样本切去上下端各 $100\alpha\%$ 数目的数据，将余下的取算术平均．当总体分布对称时，T_n 可用于估计分布的对称中心，当 n 为偶数，而 $n(1-\alpha)=2$ 或 n 为奇数，而 $n(1-\alpha)=1$ 时，相当于用样本中位数去估计分布对称中心．通常为利用样本中足够多的信息，α 应取得较小，例如 0.05 或 0.01 等．这要根据样本数目 n 的大小及究竟怀疑有多少异常值而定，也可指定两个值 α_1,α_2．样本左端切掉 $100\alpha_1\%$ 的数据，右端切掉 $100\alpha_2\%$ 的数据．

切尾的方法也可用于方差估计．

定义 14.6.2 在定义 14.6.1 的记号下，定义样本的**切尾方差**为

$$\widetilde{S}_n^2=\frac{1}{n(1-2\alpha)-1}\sum_{i=[n\alpha]}^{[n(1-\alpha)+1]}(X_{(i)}-T_n)^2,$$

其中 T_n 为**切尾均值**．

切尾均值 T_n 是总体均值 μ 的强相合和渐近正态估计．设总体分布为 $F(x)$，有

$$\sqrt{n}(T_n-\mu)\Big/\tilde{\sigma}\xrightarrow{\mathrm{d}}N(0,1),n\to\infty. \qquad (14.32)$$

其中

$$\tilde{\sigma}^2=\frac{1}{(1-2\alpha)^2}\int_{-x_\alpha}^{x_\alpha}(u-\mu)^2\mathrm{d}F(u), \qquad (14.33)$$

x_α 为 F 的上 $-\alpha$ 分位数．

$\tilde{\sigma}^2$ 可用 $\dfrac{\widetilde{S}_n^2}{(1-2\alpha)^2}$ 来估计，利用(14.31)的结论，当 n 较大时，可对均值 μ 作区间估计或假设检验．

使用切尾方法的一个问题是：究竟该切掉多长的尾，即 α 取多大是事先很难掌握的，弄不好会把正常的数据切掉．下面介绍一类

有统一数学形式的估计方法，常用于估计总体的位置参数，特别是分布的中心位置. 这类估计统称为 M 估计.

引理 14.6.3 设 $\rho(u)$ 为定义在 $(-\infty,\infty)$ 的函数，且存在 u_0，使 ρ 在 $(-\infty,u_0]$ 非增，在 $[u_0,\infty)$ 非降. 又设 F 为一连续分布函数，满足条件：

$\lambda(t)=\int_{-\infty}^{\infty}\rho(x-t)\mathrm{d}F(x)$ 对 $\forall t\in(-\infty,\infty)$ 都存在，且存在 t_0，使得 $\lambda(t_0)<\infty$. 设实数 $T_\rho(F)$ 满足

$$\lambda(T_\rho(F))=\min_{\infty<t<\infty}\lambda(t),\tag{14.34}$$

则 $T_\rho(F)$ 为 F 的一个位置参数.

例 14.6.4 在引理 14.6.3 的记号下

(1) 当 $\rho(u)=u^2$ 时，$T_\rho(F)=\mu$： F 的均值.

(2) 当 $\rho(u)=|u|$ 时，$T_\rho(F)=m$： F 的中位数.

(3) 当 $\rho(u)=\begin{cases}qu, & \text{当 } u>0\\ -pu, & \text{当 } u\leqslant 0\end{cases}$，其中 $p>0,q>0,p+q=1$，则 $T_\rho(F)=x_p$：F 的上 p 分位数.

对引理 14.6.3 所定义的位置参数 $T_\rho(F)$，自然可以考虑用 F 总体所得到样本 $X_1,\cdots,X_n$ 的经验分布函数去代替 F，而得到一个估计 $T_\rho(F_n)$，简记作 T_n. 容易验证 T_n 满足：

$$\sum_{i=1}^{n}\rho(X_i-T_n)=\min_{t}\sum_{i=1}^{n}\rho(X_i-t).\tag{14.35}$$

定义 14.6.5 在引理 14.6.3 的条件和记号下，满足(14.34)式的统计量 T_n 称为 $T_\rho(F)$ 的 **M 估计**.

例 14.6.6

(1) 当 $\rho(u)=u^2$ 时，$T_n=\overline{X}$： 样本均值.

(2) 当 $\rho(u)=|u|$ 时，$T_n=m(X)$： 样本中位数.

(3) 当 $\rho(u)=\begin{cases}(1-p)u, & \text{当 } u>0\\ -pu, & \text{当 } u\leqslant 0\end{cases}$，$p\in(0,1)$ 时，

$T_n=\xi_p$：样本上 p 分位数.

上例给出的几个 M 估计，有的是不稳健的，例如 $\overline{X}$（见本小节）. 而有些是稳健的，例如样本中位数 $m(X)$，由于 $m(X)$ 是由顺序统计量中间的一个或两个的算术平均得到的，因此即使在样本中混入少量过大或过小的异常值，也不会影响到 $m(X)$ 的值. 根据上述理由，对样本 p 分位数（作为总体 p 分位数的估计），可以说：当 p 愈接近 $\frac{1}{2}$ 时，该估计就愈稳健；反之，当 p 愈靠近 0 或 1 时，该估计就愈不稳健.

但 M 估计的重要性不在于严格按照定义 14.6.3 所给出的位置参数用相应的 M 估计去估计之. 重要的是：我们可以把 M 估计看成是一种寻找稳健估计的一般准则，用于估计总体分布（在某种意义下）的中心位置. 例如，当总体分布 F 为对称分布时，总体均值 μ 和中位数相重合，都是分布的对称中心. 这时如把总体对称中心看成是均值，用样本均值去估计之则是不稳健的，而用样本中位数去估计（对对称中心看成是中位数），则是稳健的. 一般，我们可用满足下述条件的统计量 T_n：

$$\sum_{i=1}^{n}|X_i - T_n|^r = \min_t \sum_{i=1}^{n}|X_i - t|^r \qquad (14.36)$$

去估计总体分布的对称中心，其中 $r\in[1,2]$ 为指定的常数. 这里相当于在(14.34)式中令 $\rho(u)=|u|^r$. 当 $r=2$ 时，便得到样本均值，而 $r=1$ 时便得到样本中位数. 在实用中可考虑取 r 在$[1,2]$区间中的值，例如令 $r=3/2$，其效果是既可使所得估计有较好的稳健性，又可多利用一些样本中的信息；所付出的代价是计算量增大. 因为当 $r\neq 1$ 或 2 时，(14.36)式没有一个简单形式的解，而需要使用数值计算方法（例如插值法）.

由(14.36)式定义的 M 估计方法还可用于回归分析，作为通常最小二乘法的一个补充和推广. 设 Y 为 1 维因变量，X 为 d 维自变量，$(Y_1,X_1),\cdots,(Y_n,X_n)$ 为取自总体 (Y,X) 的样本. 要用 X 的线性函数 $\beta_0+\beta_1'X$ 去回归 Y，其中 β_0 和 β_1（d 维向量）是待估参

数. M 估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 应满足方程:

$$\sum_{i=1}^{n} |Y_i - \hat{\beta}_0 - \hat{\beta}'_1 X_i|^r = \min_{\beta_0, \beta_1} \sum_{i=1}^{n} |Y_i - \beta_0 - \beta'_1 X_i|^r. \quad (14.37)$$

其中 $r \in [1,2]$. 当 $r=2$ 时,便得到通常的最小二乘解,当 $r<2$ 时,便得到 β_0 和 β_1 的一个稳健估计.

解 14.37 式比解 14.36 式更困难. 由于存在着 $d+1$ 个未知变量. 普通的插值法不能使用,而需要使用梯度法或牛顿法等计算导数的方法,同时还存在着局部极值的问题. 使用者需对上述算法有一定的了解和计算经验,才能上机计算.

15　回归分析

15.1　引　言

影响一个实验的各个变量之间存在相互联系，它们之间有两种方式：

（1）确定性关系：变量之间存在着完全确定的函数关系．写成一般式子为 $y=f(x_1,x_2,\cdots,x_m)$，并称 $x_1,\cdots,x_m$ 为自变量，y 为因变量．例如电路中的欧姆定律，设 V 表示电压，R 表示电阻，I 表示电流，欧姆定律指出

$$V = IR,$$

三个变量中如已知两个变量的值，就可以准确地算出第三个变量的值．

（2）相关关系：在许多实际问题中，由于变量之间的关系比较复杂，或者试验中误差的影响，使我们很难得到精确的数学表达式，从而使变量之间的关系具有某种不确定性．例如在炼钢时，熔毕碳与精炼时间有一定的关系，知道了熔毕碳可以大致估计出精炼时间，但不能精确地定出精炼时间．

但是，在大量重复试验之后，我们可以发现这些变量具有某种确定的概率分布，变量之间服从某种统计规律．利用统计方法研究这种规律就是回归分析．回归分析主要处理连续型随机变量之间的相关关系，利用它来建立经验公式，确定最佳生产条件、预报气象与病虫害等等．

本章中，我们要讨论下列内容：

（1）根据试验数据，确定描述变量之间的关系式．

(2) 对关系式的可信程度进行统计检验.

(3) 从影响试验指标的许多变量中,选择出影响显著的变量.

(4) 建立变量间关系的非线性模型.

(5) 改进的最小二乘估计方法——岭回归.

15.2 一元线性回归

15.2.1 回归直线的求法

一元线性回归处理的是两个变量 x 与 y 之间的线性关系,也就是通常所说的拟合直线问题.

假定,我们通过试验得到两个变量 x 与 y 的若干数据对 x_i 与 $y_i(i=1,2,\cdots,n)$,这些点 (x_i,y_i) 称为样本点,将它们描绘在直角坐标系中.假定变量 x 与 y 之间存在某种线性关系.用下面模型表示.

定义 15.2.1 $y=a+bx+\varepsilon$,

其中的 a,b 是待定常数.ε 为随机变量.上式称为**一元线性回归模型**(the simple linear regression),a,b 为**回归系数**,问题是如何确定 a,b 的值? 设

$$y_i=a+b\,x_i+\varepsilon_i\quad(i=1,2,\cdots,n),$$

是在观测点 x_i 上的观测值.如果给定了 $a.b$ 的估计值 $\hat{a},\hat{b}$,计算在 x_i 点处的因变量值,得到 $\hat{y}_i=\hat{a}+\hat{b}x_i,i=1,2,\cdots,n$. 显然 $\hat{y}_i$ 与 y_i 之间存在误差 $\delta_i=y_i-\hat{y}_i.i=1,2,\cdots,n$,根据"最小二乘"原理,系数 a,b 的估计应使误差平方和

$$\begin{aligned}Q&=\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\\&=\sum_{i=1}^{n}(y_i-a-bx_i)^2\end{aligned}\tag{15.1}$$

为最小,即选择 $\hat{a},\hat{b}$ 使得下式满足:

$$\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2 = \min_{a,b}\sum_{i=1}^{n}(y_i - a - bx_i)^2, \quad (15.2)$$

称 $\hat{a}$,$\hat{b}$ 分别为 a 与 b 的最小二乘估计.

$$\hat{y} = \hat{a} + \hat{b}x, \quad (15.3)$$

称为**一元回归方程**.

对 Q 关于 a,b 分别求一阶偏导数,并令其等于零,可求得 a,b 的最小二乘估计为

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \quad (15.4)$$

$$\hat{b} = \frac{L_{xy}}{L_{xx}}, \quad (15.5)$$

其中

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i;$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i;$$

$$L_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}x_iy_i - n\bar{x}\,\bar{y};$$

$$L_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n\,\bar{x}^2.$$

15.2.2 回归方程的显著性检验

利用式(15.4)和(15.5)我们可以求出回归方程 $\hat{y}=\hat{a}+\hat{b}x$,当知道了 x 的值时,可用 $\hat{y}$ 去预测 y 的值.但是它是否基本上符合变量 x 与 y 之间的客观规律呢?也就是问,所拟合的回归直线在多大程度上能反映出 x 与 y 之间的真实关系?为了回答这个问题,我们需要找到一个能反映变量 x 与 y 之间线性关系密切程度的统计量,以便进行统计检验.下面,我们介绍回归方程的显著性检验方法.

相关系数检验法(test of correlation coefficient)

可以证明

$$L_{yy} = Q + U. \tag{15.6}$$

其中

$$L_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2.$$

$$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

$$U = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \frac{L_{xy}^2}{L_{xx}}.$$

式中 L_{yy} 表示观察值 y_i 与均值 $\bar{y}$ 之差的平方和，它反映了数据 y_i 总的波动情况；Q 是观察值 y_i 与回归值 $\hat{y}_i$ 之差的平方和，它反映了观察值偏离回归直线 $\hat{y}=a+bx$ 的程度；U 是回归值 $\hat{y}_i$ 与平均值 $\bar{y}$ 之差的平方和，它反映了因素 x 对变量 y 线性关系的密切程度. 式(15.6)告诉我们，数据 y_i 总的波动 L_{yy} 是由两方面引起的：一方面是由于因素 x 的变化所引起的波动 U，称它为**回归平方和**；另一方面是其它的偶然因素(即随机误差)对 y 引起的波动，称 Q 为**剩余平方和.**

对于给定的一组数据 $(x_i, y_i)(i=1,2,\cdots,n)$，如果点 (x_i, y_i) 的分布很接近于一条直线，则 Q 值相对于 U 值较小；反之，若点 (x_i, y_i) 分布偏离直线，则 Q 值相对于 U 值较大.

定义15.2.2 $|r| = \sqrt{1-\dfrac{Q}{L_{yy}}}$

称 r 为变量 x 与 y 的**相关系数**(correlation coefficient). 它的大小能反映出变量 x 与 y 间的线性密切程度. 它的值界于 -1 与 1 之间. 当 Q 相对于 U 值很小时，r 接近于 $|\pm 1|$，说明变量 x 与 y 的线性关系很密切，这时的回归方程有实际意义；当 Q 相对于 U 值

很大时，r 接近于0，说明 x 与 y 之间没有密切的线性关系，这时的回归方程无实际价值.

定理15.2.3 设 ρ 是总体 (x,y) 的相关系数，当假设 H_0: $\rho=0$成立时，统计量

$$t=\frac{\sqrt{n-2}r}{\sqrt{1-r^2}}$$

服从自由度为 $n-2$的 t 分布.

对于给定的显著性水平 $\alpha(0\leqslant\alpha\leqslant1)$，以及自由度 $n-2$(n 是样品点的个数)，查 t 分布表得出相应的临界值 t_α，从而对 H_0进行检验：

当 $t>t_\alpha$ 时，否定原假设，认为 x 与 y 存在线性关系；

当 $t\leqslant t_\alpha$ 时，接受原假设，认为 x 与 y 不存在线性关系.

***F* 检验法**

如果变量 x 与 y 间无线性关系，回归方程中的回归系数 $b=0$，所以检验两个变量之间是否有线性关系，只需要检验假设:

$$H_0: b=0$$

是否成立.

定理15.2.4 在假设 H_0:$b=0$成立时，L_{yy}，U，Q 分别是自由度为 $f_y=n-1$，$f_U=1$，$f_Q=n-2$的 χ^2变量；并且 Q 与 U 相互独立，于是统计量

$$F=\frac{U/f_U}{Q/f_Q}=\frac{(n-2)U}{Q}$$

服从自由度为(f_U,f_Q)的 F 分布.

对于给定的显著性水平 α，及相应的自由度$(1,n-2)$，查 F 分布得相应的临界值 F_α，如果 $F>F_\alpha$，则否定原假设 H_0，认为 y 与 x 存在线性关系；如果 $F\leqslant F_\alpha$，则接受原假设，认为 y 与 x 间不存在线性关系.

列出方差分析表如表15.1.

表 15.1

方差来源	平方和	自由度	平均平方和	F 值	显著性
回归	$U=b^2L_{xx}$	1	U	$F=\dfrac{U}{Q/(n-2)}$ 或	
剩余	$Q=\sum_{i=1}^{n}(y_i-\hat{y})^2$	$n-2$	$Q/(n-2)$	$F=\dfrac{(n-2)r^2}{1-r^2}$	
总和	$L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$	$n-1$			

15.2.3 用回归方程进行预测

有了回归方程,我们就可以根据变量 x 的取值来预测 y 的相应值.但是,由于 x 与 y 之间是相关关系,我们自然要问:用回归方程来预测 y 时的精度如何?为此,采用类似于区间估计的方法,给定置信度$1-\alpha$,(其中$0\leqslant\alpha\leqslant1$)求出在这种置信度下预测变量 y 所取的可能值范围.

利用 y 的剩余均方差

$$S_\delta=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \tag{15.7}$$

$$=\sqrt{\frac{(1-r^2)L_{yy}}{n-2}}$$

来表示变量 y 偏离回归直线的误差.

对于任一固定的 x_0,相应的观察值 y_0将以$1-\alpha$ 的概率落在区间:

$$(\hat{y}_0-Z_{\frac{\alpha}{2}}S_\delta,\hat{y}_0+Z_{\frac{\alpha}{2}}S_\delta) \tag{15.8}$$

内.

其中 $Z_{\frac{\alpha}{2}}$ 是标准正态分布上 $\frac{\alpha}{2}$ 百分位点之值.

为了形象起见,我们在回归直线 $\hat{y}=a+bx$ 的上下作两条平行虚直线

$$
\begin{aligned}
L_1: \quad & y_1 = a + bx - Z_{\frac{\alpha}{2}} S_\delta \\
L_2: \quad & y_2 = a + bx + Z_{\frac{\alpha}{2}} S_\delta
\end{aligned}
\qquad (15.9)
$$

它表明,在全部可能出现的观察值 y_i 中,大约有 $100(1-\alpha)\%$ 的点落在这两条线 L_1 与 L_2 所夹的范围内.

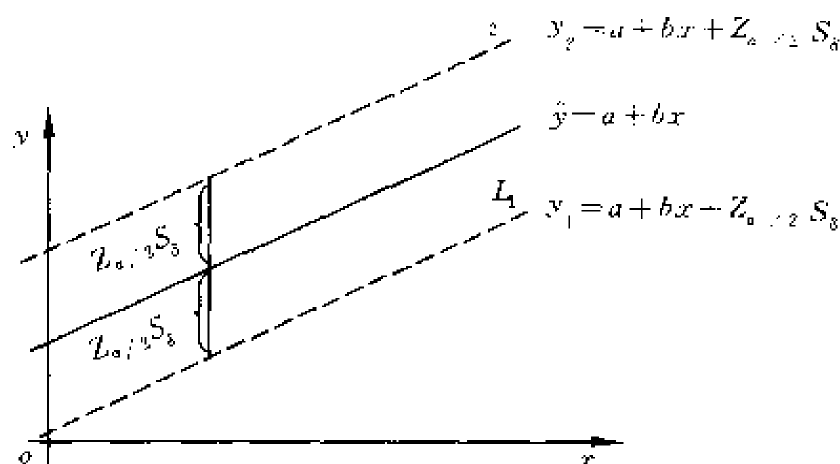


图 15.1　一元回归的预测

15.2.4　计算公式和计算步骤

(1) 为了便于计算,通常先做下述线性变换

$$
\begin{aligned}
x' &= \frac{x - c_1}{d_1} \\
y' &= \frac{y - c_2}{d_2}
\end{aligned}
\qquad (15.10)
$$

式中 c_1, c_2, d_1, d_2 为常数.

(2) 按公式(15.11)对新变量 x' 与 y' 进行计算

$$
\begin{cases}
\overline{x'} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x'_i, \\
\overline{y'} = \dfrac{1}{n}\sum\limits_{i=1}^{n} y'_i, \\
L_{x'x'} = \sum\limits_{i=1}^{n} x'^2_i - n\,\overline{x'}^2, \\
L_{x'y'} = \sum\limits_{i=1}^{n} x'_i y'_i - n\,\overline{x'}\,\overline{y'}, \\
L_{y'y'} = \sum\limits_{i=1}^{n} y'^2_i - n\,\overline{y'}^2, \\
\hat{b}' = L_{x'y'}/L_{x'x'}, \\
\hat{a}' = \overline{y'} - b'\,\overline{x'}.
\end{cases}
\tag{15.11}
$$

(3) 代回原变量，得到 x,y 的回归方程

$$
\frac{y-c_2}{d_2} = \hat{a}' + \hat{b}'\,\frac{x-c_1}{d_1}. \tag{15.12}
$$

(4) 按下述公式进行统计检验：

$$
\begin{cases}
U' = \hat{b}'^2 L_{x'x'} = \hat{b}'\,L_{x'y'}, \\
Q' = L_{y'y'} - U', \\
F' = \dfrac{(n-2)U'}{Q'}.
\end{cases}
\tag{15.13}
$$

由给定的显著性水平 α，及自由度 $(1,n-2)$ 的 F 分布表，得相应的临界值 F_α. 若 $F'>F_\alpha$，则认为**一元线性回归**模型(15.2.1)成立，若 $F'\leqslant F_\alpha$，则认为**该模型不成立**.

(5) 对回归直线进行预测

$$
S'_\delta = \sqrt{\frac{Q'}{n-2}}.
$$

$$
S_\delta = d_2 S'_\delta.
$$

据此作两条平行虚直线

$$y_1 = a + bx - Z_{\frac{\alpha}{2}}S_\delta.$$

$$y_2 = a + bx + Z_{\frac{\alpha}{2}}S_\delta.$$

可以预料 y_i 将以 $(1-\alpha)$ 的概率落在直线 y_1 与 y_2 之间。

例15.2.5 从某矿石中取得14块样品，测得 Ni 和 P_2O_5 的含量如表15.2，试分析矿石中 Ni 与 P_2O_5 的含量之间是否存在线性关系.

(1) 作线性变换

$$x' = 100(x-3),$$

$$y' = 1000(y-0.016).$$

于是将原始数据表15.2转换成新数据表15.3

表 15.2

编号	(x) Ni(%)	(y) P_2O_5(%)	编号	(x) Ni(%)	(y) P_2O_5(%)
1	0.009	4.00	8	0.014	1.70
2	0.013	3.44	9	0.016	2.92
3	0.006	3.60	10	0.014	4.80
4	0.025	1.00	11	0.016	3.28
5	0.022	2.04	12	0.012	4.16
6	0.007	4.74	13	0.020	3.35
7	0.036	0.60	14	0.018	2.20

表 15.3

编号	$x(\%)$	$y(\%)$	$x'=100(x-3)$	$y'=1000(y-0.016)$	x'^2	$x'\cdot y'$	y'^2
1	4.00	0.009	100	7	10000	−700	49
2	3.44	0.013	44	−3	1936	−132	9
3	3.60	0.006	60	−10	3600	−600	100
4	1.00	0.025	−200	9	40000	−1800	81
5	2.04	0.022	−96	6	9216	−576	36
6	4.74	0.007	174	−9	30276	−1566	81
7	0.60	0.036	−200	20	57600	−4800	400
8	1.70	0.014	−130	−2	16900	260	4
9	2.92	0.016	−8	0	64	0	0
10	4.80	0.014	180	−2	32400	−360	4
11	3.28	0.016	28	0	784	0	0
12	4.16	0.012	116	−4	13456	−464	16
13	3.35	0.020	35	4	1225	140	16
14	2.20	0.018	−80	2	6400	−160	4
Σ			−17	4	223857	−10758	800

(2) 按公式(15.11)进行计算

$$\overline{x'} = \frac{1}{14}(-17) = -1.214,$$

$$\overline{y'} = \frac{1}{14}(4) = 0.286,$$

$$L_{x'x'} = 223857 - \frac{1}{14}(-17)^2 = 223836,$$

$$L_{x'y'} = -10758 - \frac{1}{14}(-17)(4) = -10753,$$

$$L_{y'y'} = 800 - \frac{1}{14}(4)^2 = 799,$$

$$b' = \frac{L_{x'y'}}{L_{x'x'}} = -\frac{10753}{223836} = -0.048,$$

$$a' = \overline{y'} - b'\,\overline{x'} = 0.286 - (-0.048)(-1.214) = 0.228.$$

(3) 代回原变量,得 x 与 y 的回归方程

$$1000(y - 0.016) = 0.228 - 0.048 \times 100(x - 3).$$

得
$$\hat{y} = 0.0306 - 0.0048x.$$

(4) 进行统计检验

$$U' = b'L_{x'y'} = (-0.048)(-10753) = 516.144,$$

$$Q' = L_{y'y'} - U' = 799 - 516.144 = 282.856,$$

$$F' = \frac{(n-2)U'}{Q'} = \frac{12(516.144)}{282.856} = 21.96.$$

由 $\alpha = 0.05$ 查 $F_{0.05}(1,12)$ 得 $F_{0.05} = 9.93$,现在 $F = 21.96 > F_{\alpha} = 9.93$,故可认为 x 与 y 之间存在直线关系.

(5) 对回归直线进行预测:

$$S'_{\delta} = \sqrt{\frac{Q'}{n-2}} = \sqrt{\frac{282.856}{12}} = 4.9,$$

$$S_{\delta} = d_2 S'_{\delta} = \frac{1}{1000} S'_{\delta} = 0.049.$$

对于给定的置信度 $1-\alpha = 1-5\% = 95\%$,查出 $Z_{0.025} = 1.96$.

$$L_1:\ y_1 = a + bx - 1.96S_{\delta} = 0.0210 - 0.0048x,$$

$$L_2:\ y_2 = a + bx + 1.96S_{\delta} = 0.0402 - 0.0048x.$$

于是可以预料 Ni 将以95%的概率落在直线 L_1 与 L_2 之间.

15.3 多元线性回归

在许多问题中,影响试验结果的变量不止一个,例如有 p 个:$x_1, x_2, \cdots, x_p$. 研究因变量 y 与自变量 $x_1, x_2, \cdots, x_p$ 之间的定量关系是多元回归分析的任务. 在多元回归分析中,我们只讨论多元线

性回归(multiple linear regresson)问题.从原理上讲,多元线性回归分析与一元线性回归分析相同,但计算比较复杂.

15.3.1 多元回归方程的求法

假设有 p 个自变量 $x_1,x_2,\cdots,x_p$,而 y 是表示实验指标的因变量.现在共进行了 n 次试验,试验结果是:

$$(y_i;x_{i1},x_{i2},\cdots,x_{ip}) \qquad i=1,2,\cdots,n.$$

假定因变量 y 与 p 个自变量 $x_1,\cdots,x_p$ 间存在线性关系,并可用下述数学模型表达:

$$y=\beta_0+\beta_1x_1+\cdots+\beta_px_p+\varepsilon \tag{15.14}$$

将 n 次试验结果代入,可得如下的结构式

$$\begin{cases} y_1=\beta_0+\beta_1x_{11}+\beta_2x_{12}+\cdots+\beta_px_{1p}+\varepsilon_1, \\ y_2=\beta_0+\beta_1x_{21}+\beta_2x_{22}+\cdots+\beta_px_{2p}+\varepsilon_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n=\beta_0+\beta_1x_{n1}+\beta_2x_{n2}+\cdots+\beta_px_{np}+\varepsilon_n. \end{cases} \tag{15.15}$$

其中 $\beta_0,\beta_1,\cdots,\beta_p$ 是待估计的参数称为回归系数,$\varepsilon_1,\cdots,\varepsilon_n$ 是 n 个相互独立且服从同一正态分布 $N(0,\sigma^2)$ 的随机变量.

引用下面矩阵符号:

$$\boldsymbol{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \boldsymbol{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\boldsymbol{\beta}=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \qquad \boldsymbol{\varepsilon}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

由此,(15.15)可以写成矩阵形式:

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{15.16}$$

我们的目的是估计参数 $\boldsymbol{\beta}$. 设 $b_0, b_1, \cdots, b_p$ 分别是 $\beta_0, \beta_1, \cdots, \beta_p$ 的最小二乘估计，于是回归方程为

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p \tag{15.17}$$

根据最小二乘原理，回归系数 $b_0, b_1, \cdots, b_p$ 应使得全部观察值 y_i 与回归值 $\hat{y}_i$ 间的离差平方和 Q

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

为最小.

将 Q 对 $b_0, b_1, \cdots, b_p$ 求偏导数，令诸偏导数为零，并将回归方程(15.17)在各次试验下的回归值 $\hat{y}_i (i = 1, \cdots, n)$ 代入，可得下面的正规方程(normal equation)

$$\boldsymbol{A} \cdot \boldsymbol{b} = \boldsymbol{B} \tag{15.18}$$

其中

$$\boldsymbol{A} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1} x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1} x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{i1} x_{ip} & \sum_{i=1}^{n} x_{i2} x_{ip} & \cdots & \sum_{i=1}^{n} x_{ip}^2 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ip} y_i \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}.$$

不难证明

$$\boldsymbol{A} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}. \tag{15.19}$$

$$\boldsymbol{B} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}. \tag{15.20}$$

于是正规方程(15.18)可以写成另一形式：

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{b} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}). \tag{15.21}$$

在 $\boldsymbol{A}$ 矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 满秩的情况下，存在逆矩阵 $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$，因而得

$$\boldsymbol{b} = \boldsymbol{A}^{-1}\boldsymbol{B} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}. \tag{15.22}$$

式(15.22)是模型(15.15)中参数 β 的最小二乘估计.

15.3.2 计算公式和计算步骤

在实际计算时，为了方便常采用如下形式的回归模型：

$$\begin{aligned} y_i = \mu_0 &+ \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) \\ &+ \cdots + \beta_p(x_{ip} - \overline{x}_p) + \varepsilon_i, \quad i = 1,2,\cdots,n. \end{aligned} \tag{15.23}$$

其中的 $\overline{x}_i = \dfrac{1}{n}\sum\limits_{k=1}^{n} x_{ki} (i = 1,2,\cdots,p)$.

显然有 $\mu_0 - \sum\limits_{i=1}^{p}\beta_i \overline{x_i} = b_0$.

这时，相应的设计矩阵 $\boldsymbol{X}$，矩阵 $\boldsymbol{B}$，以及正规方程系数矩阵 $\boldsymbol{A}$ 分别为

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} - \overline{x}_1 & x_{12} - \overline{x}_2 & \cdots & x_{1p} - \overline{x}_p \\ 1 & x_{21} - \overline{x}_1 & x_{22} - \overline{x}_2 & \cdots & x_{2p} - \overline{x}_p \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} - \overline{x}_1 & x_{n2} - \overline{x}_2 & \cdots & x_{np} - \overline{x}_p \end{bmatrix} \tag{15.24}$$

$$\boldsymbol{B} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ l_{1y} \\ \vdots \\ l_{py} \end{bmatrix}.$$

$$A = \begin{bmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & l_{11} & l_{12} & \cdots & l_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & L \end{bmatrix}, \tag{15.25}$$

其中

$$l_{ij} = \sum_{k=1}^{n} x_{ki} x_{kj} - n\bar{x}_i \bar{x}_j, \quad i, j = 1, 2, \cdots, p. \tag{15.26}$$

$$l_{jy} = \sum_{i=1}^{n} x_{ij} y_i - n\bar{x}_j \bar{y}, \quad j = 1, 2, \cdots, p. \tag{15.27}$$

利用求(15.18)相同的方法,可得下面的正规方程

$$\boldsymbol{A}\boldsymbol{b} = \boldsymbol{B}, \tag{15.28}$$

或

$$\boldsymbol{b} = \boldsymbol{A}^{-1}\boldsymbol{B}. \tag{15.29}$$

即

$$\boldsymbol{b} = \begin{bmatrix} \mu_0 \\ \mathrm{b}_1 \\ \vdots \\ \mathrm{b}_p \end{bmatrix} = \begin{bmatrix} 1/n & 0 \\ & \\ 0 & L^{-1} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ l_{1y} \\ \vdots \\ l_{py} \end{bmatrix}, \tag{15.30}$$

得

$$\mu_0 = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$$

$$\begin{bmatrix} \mathrm{b}_1 \\ \mathrm{b}_2 \\ \vdots \\ \mathrm{b}_p \end{bmatrix} = L^{-1} \begin{bmatrix} l_{1y} \\ l_{2y} \\ \vdots \\ l_{py} \end{bmatrix} \triangleq L^{-1} \begin{bmatrix} \mathrm{B}_1 \\ \mathrm{B}_2 \\ \vdots \\ \mathrm{B}_p \end{bmatrix}, \tag{15.31}$$

其中

$$\mathrm{B}_i = l_{iy}.$$

$$
\boldsymbol{L} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1p} \\ l_{21} & l_{22} & \cdots & l_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix},
$$

则回归系数为

$$
\begin{cases} b_0 = \overline{y} - \sum_{i=1}^{p} b_i \overline{x}_i \\ \boldsymbol{b} = \begin{bmatrix} \mathrm{b}_1 \\ \mathrm{b}_2 \\ \vdots \\ \mathrm{b}_p \end{bmatrix} = \boldsymbol{L}^{-1} \begin{bmatrix} l_{1y} \\ l_{2y} \\ \vdots \\ l_{py} \end{bmatrix} \triangleq \boldsymbol{L}^{-1} \begin{bmatrix} \mathrm{B}_1 \\ \mathrm{B}_2 \\ \vdots \\ \mathrm{B}_p \end{bmatrix}. \end{cases}
$$

其中 $\mathrm{B}_i = l_{iy}$.

例15.3.1 平炉炼钢过程中,由于矿石及炉气的氧化作用,铁水的总含碳量在不断降低,一炉钢在冶炼初期(熔化期)中总的去碳量 y,与所加的两种矿石(天然矿石与烧结矿石)的量 x_1,x_2及溶化时间 x_3有关,得试验数据表15.4.

(1) 为了求出总去碳量 y 对变量 x_1,x_2,x_3的线性回归方程,我们利用回归模型(15.23),这时 $p=3$:

$$
y_i = \mu + \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \beta_3(x_{i3} - \overline{x}_3) + \varepsilon_i,
$$
$$
i = 1, 2, \cdots, 49.
$$

(2) 计算各变量的总和,算术平均值,交叉乘积和 l_{ij},以及 l_{iy}:

$n=49$,

$$\sum_{i=1}^{49} y_i = 224.5169, \qquad \overline{y} = 4.582,$$

$$\sum_{i=1}^{49} x_{i1} = 259, \qquad \overline{x}_1 = 5.286,$$

$$\sum_{i=1}^{49} x_{i2} = 578, \qquad \overline{x}_2 = 11.796,$$

表 15.4

编号	Y (t)	X_1 (槽)	X_2 (槽)	X_3 (5min)	编号	Y (t)	X_1 (槽)	X_2 (槽)	X_3 (5min)
1	4.3302	2	18	50	26	2.7066	9	6	39
2	3.6458	7	9	40	27	5.6314	12	5	51
3	4.4830	5	14	46	28	5.8152	6	13	41
4	5.5468	12	3	43	29	5.1302	12	7	47
5	5.4970	1	20	64	30	5.3910	0	24	61
6	3.1125	3	12	40	31	4.4583	5	12	37
7	5.1182	3	17	64	32	4.6569	4	15	49
8	3.8759	6	5	39	33	4.5212	0	20	45
9	4.6700	7	8	37	34	4.8650	6	16	42
10	4.9536	0	23	55	35	5.3566	4	17	48
11	5.0060	3	16	60	36	4.6098	10	4	48
12	5.2701	0	18	49	37	2.3815	4	14	36
13	5.3772	8	4	50	38	3.8746	5	13	36
14	5.4849	6	14	51	39	4.5919	9	18	51
15	4.5960	0	21	51	40	5.1588	6	13	54
16	5.6645	3	14	51	41	5.4373	5	18	100
17	6.0795	7	12	56	42	3.9960	5	11	44
18	3.2194	16	0	48	43	4.3970	8	6	63
19	5.8076	6	16	45	44	4.0622	2	13	55
20	4.7306	0	15	52	45	2.2905	7	8	50
21	4.6805	9	0	40	46	4.7115	4	10	45
22	3.1272	4	6	32	47	4.5310	10	5	40
23	2.6104	0	17	47	48	5.3637	3	17	64
24	3.7174	9	0	44	49	6.0771	4	15	72
25	3.8946	2	6	39					

$$\sum_{i=1}^{49} x_{i3} = 2411, \qquad \bar{x}_3 = 49.204,$$

$$\sum_{i=1}^{49} x_{i1}^2 = 2031, \qquad \sum_{i=1}^{49} x_{i2}^2 = 8572,$$

$$\sum_{i=1}^{49} x_{i3}^2 = 124879, \qquad \sum_{i=1}^{49} x_{i1}x_{i2} = 2137,$$

$$\sum_{i=1}^{49} x_{i1}x_{i3} = 12355, \qquad \sum_{i=1}^{49} x_{i2}x_{i3} = 29216,$$

$$\sum_{i=1}^{49} x_{i1}y_i = 1180.30, \qquad \sum_{i=1}^{49} x_{i2}y_i = 2717.51,$$

$$\sum_{i=1}^{49} x_{i3}y_i = 11292.72.$$

由此得

$$l_{11} = \sum_{i=1}^{49} x_{i1}^2 - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i1}\right)^2 = 662.000,$$

$$l_{22} = \sum_{i=1}^{49} x_{i2}^2 - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i2}\right)^2 = 1753.959,$$

$$l_{33} = \sum_{i=1}^{49} x_{i3}^2 - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i3}\right)^2 = 6247.959,$$

$$l_{21} = l_{12} = \sum_{i=1}^{49} x_{i1}x_{i2} - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i1}\right)\left(\sum_{i=1}^{49} x_{i2}\right) = -918.143,$$

$$l_{13} = l_{31} = \sum_{i=1}^{49} x_{i1}x_{i3} - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i1}\right)\left(\sum_{i=1}^{49} x_{i3}\right) = -388.857,$$

$$l_{32} = l_{23} = \sum_{i=1}^{49} x_{i2}x_{i3} - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i2}\right)\left(\sum_{i=1}^{49} x_{i3}\right) = 776.041,$$

$$l_{1y} = \sum_{i=1}^{49} x_{i1}y_i - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i1}\right)\left(\sum_{i=1}^{49} y_i\right) = -6.433,$$

$$l_{2y} = \sum_{i=1}^{49} x_{i2}y_i - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i2}\right)\left(\sum_{i=1}^{49} y_i\right) = 69.130,$$

$$l_{3y} = \sum_{i=1}^{49} x_{i3}y_i - \frac{1}{n}\left(\sum_{i=1}^{49} x_{i3}\right)\left(\sum_{i=1}^{49} y_i\right) = 245.571.$$

(3) 计算系数矩阵 $\boldsymbol{A}$,常数项矩阵 $\boldsymbol{B}$,以及逆矩阵 $\boldsymbol{A}^{-1}$.

$$\boldsymbol{A}=\begin{bmatrix} n & 0 \\ 0 & L \end{bmatrix}=\begin{bmatrix} 49 & 0 & 0 & 0 \\ 0 & 662.000 & -918.143 & -388.857 \\ 0 & -918.143 & 1753.959 & 776.041 \\ 0 & -388.857 & 776.041 & 6247.959 \end{bmatrix},$$

$$\boldsymbol{B}=\begin{bmatrix} \sum_{i=1}^{49} y_i \\ l_{1y} \\ \vdots \\ l_{3y} \end{bmatrix}=\begin{bmatrix} 224.5196 \\ -6.433 \\ 69.130 \\ 245.571 \end{bmatrix},$$

$$\boldsymbol{A}^{-1}=\begin{bmatrix} \frac{1}{49} & 0 & 0 & 0 \\ 0 & 0.005515 & 0.002894 & -0.00001623 \\ 0 & 0.002984 & 0.002122 & -0.00008345 \\ 0 & -0.00001623 & -0.00008345 & 0.0001694 \end{bmatrix}.$$

(4) 计算回归系数和回归方程

$$\boldsymbol{b}=\begin{bmatrix} \mu_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}=\boldsymbol{A}^{-1}\boldsymbol{B}=\begin{bmatrix} 4.582 \\ 0.1604 \\ 0.1076 \\ 0.0359 \end{bmatrix}.$$

由此得回归方程

$$\hat{y}=4.582+0.1604(x_1-5.286)+0.1076(x_2-11.769)$$
$$+0.0359(x_3-49.204),$$

即 $$\hat{y}=0.7014+0.1604x_1+0.1076x_2+0.0359x_3.$$

15.3.3 回归方程的显著性检验

和一元线性回归情形类似,对于给定的观测数据:

$$y_i, x_{i1}, x_{i2}, \cdots, x_{ip} \quad (i=1,2,\cdots,n),$$

总可以按最小二乘原理配一个线性回归方程.但这样配置的线性回归方程是否有实际意义,还须进行统计检验.

同一元线性回归的情形一样可以证明

(1) 总的离差平方和 L_{yy} 可以分成 Q 与 U 两部分

$$L_{yy} = Q + U, \tag{15.32}$$

其中

$$L_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$U = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2,$$

$$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

(2) 在假设

$$H_0: b_1 = b_2 = \cdots = b_p = 0 \tag{15.33}$$

成立时,L_{yy},U,Q 分别是自由度 $f_y=n-1$,$f_U=p$,$f_Q=n-p-1$ 的 χ^2 变量,并且 Q 与 U 相互独立.

(3) 在假设 H_0 成立时,统计量:

$$F = \frac{U/p}{Q/(n-p-1)} = \frac{(n-p-1)U}{pQ} \tag{15.34}$$

服从自由度为$(p,n-p-1)$的 F 分布.这样,我们可按下述步骤进行统计检验:

(1) 按方差分析表15.5进行计算.

表 15.5

来源	平方和	自由度	均方	F
回归	$U = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}) = \sum_{j=1}^{p} b_j B_j$	p	U/p	$\frac{U/p}{Q/(n-p-1)}$
剩余	$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i) = L_{yy} - U$	$n-p-1$	$Q/(n-p-1)$	
总计	$L_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$	$n-1$		

(2) 根据自由度$(p,n-p-1)$，显著性水平α查F分布表，得相应的临界值F_α，若$F>F_\alpha$，则否定原假设H_0，认为y与$x_1,x_2,\cdots,x_p$之间存在线性关系，若$F\leqslant F_\alpha$，则接受原假设H_0，认为y与$x_1,x_2,\cdots,x_p$之间不存在线性关系.

例15.3.2 在例15.3.1中，据表15.4可以求得

$$L_{yy}=\sum_{i=1}^{49}y_i^2-\frac{1}{n}\Big(\sum_{i=1}^{n}y_i\Big)^2=44.905,$$

$$U=\sum_{j=1}^{3}b_jB_j=0.1604\times(-6.433)+0.1076\times 69.130$$
$$+0.0359\times 245.571=15.221,$$
$$Q=L_{yy}-U=44.905-15.221=29.684,$$
$$f_U=3,$$
$$f_Q=49-3-1=45,$$
$$U/f_U=5.074,$$
$$Q/f_Q=0.660,$$
$$F=\frac{5.074}{0.660}=7.69.$$

若给定$\alpha=0.01$，由自由度(3,45)查F分布表得$F_{0.01}(3,45)=4.25$.

现在$F=7.69>F_{0.01}(3,45)=4.25$. 检验结果表明例15.3.1中的$y$与$x_1,x_2,x_3$之间存在线性关系.

15.3.4 用回归方程进行预测

在前面，我们建立了多元线性回归方程：

$$\hat{y}=b_0+b_1x_1+b_2x_2+\cdots+b_px_p.$$

在对它进行了统计检验之后，便能够利用它进行预测.

与一元线性回归时类似，用

$$S_\delta = \sqrt{\frac{Q}{n-p-1}}$$

近似地表示 y 偏离回归平面的误差. 于是,可以预测在各变量 $x_1,\cdots,x_p$ 取固定值时的观测值,将以 $1-\alpha$ 的概率落在下述区间内:

$$\left(\hat{y}_0 - Z_{\frac{\alpha}{2}}S_\delta, \hat{y}_0 + Z_{\frac{\alpha}{2}}S_\delta\right).$$

15.3.5 偏回归系数的显著性检验

在(15.3.3)中介绍了回归方程的显著性检验,利用(15.34)计算 F,若 $F>F_\alpha$,则认为回归效果显著. 但此时并不意味着所引入的每个自变量 x 都在回归方程中起着重要作用,为此还需要检验各个自变量对 y 是否有显著作用.

定义15.3.3 称 $V_i=\frac{b_i^2}{c_{ii}}$ 为 y 对 x_i 的**偏回归平方和**. 其中 c_{ii} 是矩阵 $\boldsymbol{L}$ 的逆矩阵 $\boldsymbol{L}^{-1}$ 中对角线上第 i 个元素. 统计量

$$F_i = \frac{V_i}{Q/(n-p-1)}$$

在 $H_0:b_i=0$ 成立时,服从自由度为 $(1,n-p-1)$ 的 F 分布. 对于给定显著性水平 α,查自由度为 $(1,n-p-1)$ 的 F 表得临界值 F_α. 若 $F_i>F_\alpha$,则说明第 i 个自变量 x_i 对 y 影响显著,这变量应保留在回归方程中. 若经检验是不显著的变量,则可根据实际情况把偏回归平方和最小的那个变量去掉. 比如该变量是 x_j,由回归方程中去掉一个变量 x_j 后,缺少一个变量 x_j 的新回归方程的系数 b_i^* 仍需要重新计算,可用公式

$$b_i^* = b_i - \frac{c_{ij}}{c_{jj}}b_j, \quad (i \neq j)$$

其中 c_{ij} 是 $\boldsymbol{L}^{-1}$ 中第 (i,j) 个元素,c_{jj} 是 $\boldsymbol{L}^{-1}$ 中第 (j,j) 个元素,从而得到新回归方程.

15.4 逐步回归

15.4.1 逐步回归的基本思想

在实际问题中，可能有许多自变量对因变量 y 有或大或小的影响. 考虑到随机误差的存在，那些对因变量影响较小的自变量在回归方程中起的作用可能是不真实的，或者只起到分散注意力的作用. 为使回归方程能够真实地反映自变量与应变量之间的关系，应该在最终选定的回归方程中只包含那些对因变量有显著影响的自变量. **逐步回归**(stepwrse regresson)方法就是诸多选择变量方法中效果较好且计算量不大的一种方法，有着广泛的应用.

逐步回归的基本思想是从众多的自变量中，根据这些变量各自对回归方程影响的大小，逐次地选入到回归方程中，在这个过程中，先前被选入回归方程的变量，有些由于其后新引入的变量而失去了重要性，这时就应从回归方程中将它们淘汰掉. 持续上述过程，直到回归方程不再有可淘汰的变量，也没有再可引入的变量时为止. 最后所得的结果就是选定的回归方程.

现在的主要问题是

(1) 如何度量某个变量在回归方程中作出的贡献；

(2) 如何检验某个变量的作用是否是显著的.

由(15.32)知

$$L_{yy} = Q + U$$

即总的离差平方和 L_{yy} 等于回归平方和 U 与剩余平方和 Q 之和. 对于给定的试验数据，L_{yy} 是固定的. 由 p 个变量的回归方程给出的回归平方和为：

$$U = \sum_{j=1}^{p} b_j B_j = L_{yy} - Q \qquad (15.35)$$

B_j 见(15.31). 如若从这 p 个变量中去掉一个变量 x_i，重新计算余

下的 $p-1$ 个变量的回归系数，记作 b_j'，然后算出相应的回归平方和

$$U' = \sum_{\substack{j=1 \\ j\neq i}}^{p} b_j' B_j = L_{yy} - Q' \tag{15.36}$$

这样一来，就可以求出某个变量 x_i 在 p 个变量的回归方程中作出的贡献，记作 V_i，则有

$$V_i = U - U' = Q' - Q \tag{15.37}$$

称 V_i 为偏回归平方和. 可以证明

$$V_i = \frac{b_i^2}{c_{ii}} \tag{15.38}$$

V_i 愈大，则相应的 x_i 愈重要. 式中的 b_i 是对应于 x_i 的回归系数；c_{ii}是正规方程系数矩阵 L 的逆矩阵 L^{-1}主对角线上的第 i 个元素.

显然，偏回归平方和 V_i 愈大的 x_i，对 y 有重要的影响. 但是 V_i 大到什么程度才能算作是显著的呢?这就需要进行统计检验.

统计量

$$F_i = \frac{V_i/1}{Q/(n-p-1)} \tag{15.39}$$

在 $\beta_i=0$的假设下，服从自由度为$(1,n-p-1)$的 F 分布. 利用它可以检验 x_i 的显著性. 如果 $F<F_\alpha$，则 x_i 的作用不显著，可以从回归方程中剔除；如果 $F\geqslant F_\alpha$，则 x_i 不应剔除. 是否剔除一个变量，要看它对回归方程的贡献是否显著.

15.4.2 逐步回归的计算步骤

计算可以分成以下几个步骤.

步骤1：计算均值，标准差，离差矩阵以及相关矩阵：

(1) 求均值

$$\overline{x_i} = \frac{1}{n}\sum_{k=1}^{n} x_{ki}, i = 1,2,\cdots,p \tag{15.40}$$

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_i$$

（2）求离差矩阵，其元素 s_{ij} 为：

$$s_{ij} = s_{ji} = \sum_{k=1}^{n}(x_{ki} - \overline{x_i})(x_{kj} - \overline{x_j}),(i,j = 1,2,\cdots,p) \tag{15.41}$$

为了后续计算，在离差矩阵上增加一个新列，即 $p+1$列，由下面公式给出

$$s_{iy} = \sum_{k=1}^{n}(x_{ki} - \overline{x_i})(y_k - \bar{y}),(i = 1,2,\cdots,p) \tag{15.42}$$

根据离差矩阵的对角线元素，利用下述公式可以算各个标准差：

$$\sqrt{s_{ii}/(n-1)}. \tag{15.43}$$

可得到表15.6.

表　15.6

		x_1	x_2	$\cdots$	x_p	y(第 $p+1$列)
均值		$\frac{1}{n}\sum_{k=1}^{n} x_{k1}$			$\frac{1}{n}\sum_{k=1}^{n} x_{kp}$	$\frac{1}{n}\sum_{k=1}^{n} y_i$
标准差 $\sqrt{s_{ii}/(n-1)}$						
s_{ij}	x_1	s_{11}	s_{12}	$\cdots$	s_{1p}	s_{1y}
	x_2	s_{21}	s_{22}	$\cdots$	s_{2p}	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	x_p	s_{p1}	s_{p2}	$\cdots$	s_{pp}	s_{py}
s_{yy}						

（3）为计算方便起见，作代换

$$r_{ij}=\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}, \quad (i=1,2,\cdots,p;j=1,2,\cdots,p,y)$$

(15.44)

将 p 行，$p+1$列的扩展离差矩阵变换为(15.44)相关系数矩阵. 逐步计算均以此矩阵为依据.

步骤2：逐步计算

假设已计算到了 l 步(包括第 $l=0$步)，其对应的相关系数矩阵元素用上标 l 标记为 $r_{ij}^{(l)}(i=1,2,\cdots,p;j=1,2,\cdots,p,y)$. 开始下述计算：

(1) 求出每一个变量对回归方程的贡献：

如果在第 $l+1$步要从回归方程中淘汰变量 x_i，可以证明损失的贡献为

$$\widetilde{V}_i^{(l)}=(r_{iy}^{(l)})^2/r_{ii}^{(l)} \quad (15.45)$$

如果在第 $l+1$步，将某个变量 x_k 引入到回归方程中，可以证明增加的贡献为

$$\widetilde{V}_k^{(l+1)}=(r_{ky}^{(l)})^2/r_{kk}^{(l)} \quad (15.46)$$

由公式(15.45)～(15.46)看出，由相关矩阵第 $p+1$列与相应对角线元素求出的贡献$(r_{iy}^{(l)})^2/r_{ii}^{(l)}$，既可看成是已选变量被淘汰后损失的贡献，也可看成是新引入变量所增加的贡献.

(2) 在诸被引入的变量中，求出可以淘汰的变量.

为此，在各已引入的变量中求出具有最小贡献的那一个变量，然后对它进行统计检验，确认这个候选淘汰对象是否达到被淘汰的标准.

计算贡献最小的变量：

$$\widetilde{V}_k^{(l)}=\min\{\widetilde{V}_i^{(l)}\mid i\text{ 是各已引入变量的编号}\} \quad (15.47)$$

计算相应的 F 值：

$$F=(n-r-1)\widetilde{V}_k^{(l)}/\widetilde{Q}^{(l)} \quad (15.48)$$

其中 l 表示进行到第 l 步计算

n 是观测次数

r 是欲在第 $l+1$ 步对已引入的 r 个自变量中剔除某个自变量,如 x_k,$\tilde{Q}^{(l)}$是第 l 步的残差平方和,$\tilde{Q}^{(l)}=r_{yy}^{(l)}$可以利用下述公式逐步算出:

$$\tilde{Q}^{(0)}=1 \tag{15.49}$$
$$\tilde{Q}^{(l)}=\tilde{Q}^{(l-1)}-\tilde{V}_k^{(l)}.$$

如果对计算值 F 有:$F\leqslant F_a$,则将变量 x_k 从回归方程中淘汰出去. 然后开始消去运算.

(3) 如果对计算值 F 有:$F>F_a$,则 x_k 不应从回归中剔除,应考虑从未引入的各变量中选出贡献最大的变量,计算:

$$\tilde{V}_k^{(l+1)}=\max\{\tilde{V}_i^{(l)}|i\text{ 是各未选变量的编号}\} \tag{15.50}$$

求相应的 F 值:

$$F=(n-r-2)\tilde{V}_k^{(l+1)}/(\tilde{Q}^{(l)}-\tilde{V}_k^{(l+1)}) \tag{15.51}$$

如果 $F>F_a$,则把 x_k 引进到回归方程中,然后开始消去运算. 如果 $F<F_a$ 表示 x_k 不能被引入回归方程.

(4) 消去运算. 对需要淘汰或引进的变量 x_k 进行一次消去运算,即按下述公式对相关系数矩阵进行计算:

$$r_{ij}^{(l+1)}=\begin{cases}r_{kj}^{(l)}/r_{kk}^{(l)} & (i=k,j\neq k),\\ r_{ij}^{(l)}-r_{ik}^{(l)}r_{kj}^{(l)}/r_{kk}^{(l)} & (i\neq k,j\neq k),\\ 1/r_{kk}^{(l)} & (i=k,j=k),\\ -r_{ik}^{(l)}/r_{kk}^{(l)} & (i\neq k,j=k).\end{cases} \tag{15.52}$$

然后重复(15.47)~(15.52)直至在既不能剔除,也无法再被重新引入下,逐步计算过程可以结束. 转入步骤3的计算.

步骤3　求回归方程及相应的量.

设在 l 步后停止逐步计算,这时引进了 r 个变量,对各引入变量 x_i 按下面公式计算回归系数:

$$b_i^{(l)}=r_{iy}^{(l)}\sqrt{s_{yy}}/\sqrt{s_{ii}}=r_{iy}^{(l)}\frac{s_y}{s_i}$$

以及 $i=1,2,\cdots,r$

常数项 $b_0^{(l)}=\overline{y}-\sum\limits_{\text{被引入的各}x_i} b_i^{(l)}\overline{x}_i.$

其中 $s_y=\sqrt{\frac{1}{n-1}\sum\limits_{k=1}^{n}(y_k-\overline{y})^2}=\sqrt{\frac{1}{n-1}s_{yy}}$

$$s_i=\sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_{ki}-\overline{x}_i)^2}=\sqrt{\frac{1}{n-1}s_{ii}}.$$

残差平方和 $Q^{(l)}=s_{yy}r_{yy}^{(l)}.$

回归平方和 $U^{(l)}=s_{yy}-Q^{(l)}.$

复相关系数 $R=\sqrt{1-\frac{Q^{(l)}}{s_{yy}}}=\sqrt{1-r_{yy}^{(l)}}.$

其中复相关系数 R 与一元线性回归中 r(15.2.2)的含义相似,表示因变量 y 对于自变量 $x_1,x_2,\cdots,x_p$ 的总体线性相关性,它的值在0和1之间.当 $R=1$时,表示 y 与诸自变量之间有严格的线性关系;而当 $R=0$时,表示 y 与诸自变量之间完全没有线性关系.当 R 接近1时(比如 $R\geqslant 0.85$),说明用诸自变量的线性函数去刻画因变量是适宜的;反之,当 R 接近于零时,说明用自变量的线性函数去刻画因变量是不适宜的,应该考虑非线性回归方程,或引入新的变量.

15.5 回归系数的有偏估计——岭回归

在线性回归模型(见式15.16)

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$$

之下,利用最小二乘原理,可以得到 $\boldsymbol{\beta}$ 的最小二乘估计

$$\boldsymbol{b}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

可以证明最小二乘估计为方差最小的线性无偏估计,在正态分布下是方差最小无偏估计.也就是说,当限制在无偏估计类中

时，最小二乘估计有优良性质.

另一方面，在处理回归分析问题中不断积累的经验显示，在某些情况下，利用最小二乘估计 $\boldsymbol{b}$，结果并不理想. 究其原因知与设计矩阵 $\boldsymbol{X}$ 的特性有关. 当系数矩阵 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 呈现“病态”，即 $\boldsymbol{A}$ 的行列式 $|\boldsymbol{A}|$ 很小(接近于零)时，最小二乘估计的性能变得很坏.

可以证明，在线性回归模型

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{15.53}$$

以及

$$E(\boldsymbol{\varepsilon})=0 \tag{15.54}$$

$$\operatorname{var}(\boldsymbol{\varepsilon})=\sigma^2 I(I\text{ 是单位矩阵})$$

的假定之下，有

$$E(\|\boldsymbol{b}-\boldsymbol{\beta}\|^2)=\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} \tag{15.55}$$

$$\operatorname{var}(\|\boldsymbol{b}-\boldsymbol{\beta}\|^2)=2\sigma^4\sum_{i=1}^{p}\frac{1}{\lambda_i^2} \tag{15.56}$$

其中 λ_i 是系数矩阵 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 的特征根.

由(15.55)～(15.56)看出：当 $\boldsymbol{A}$ 呈现“病态”时，$E(\|\boldsymbol{b}-\boldsymbol{\beta}\|^2)$ 与 $\operatorname{var}(\|\boldsymbol{b}-\boldsymbol{\beta}\|^2)$ 的值都很大，这表明 b 与 β 之间的偏差的变化很大，而且不稳定. 因此，在矩阵 $\boldsymbol{A}$“病态”的情况下，不能认为最小二乘估计 $\boldsymbol{b}$ 是 $\boldsymbol{\beta}$ 的良好估计了.

当 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 是奇异矩阵时(即 $|A|=0$)，$\boldsymbol{X}$ 的列向量线性相关. 如果 A 接近奇异，即 $|\boldsymbol{A}|$ 接近零，那么 $\boldsymbol{X}$ 的列向量接近线性相关，也就是说，在自变量之间存在近似线性关系，通常把这种近似线性关系称为**复共线关系**. 当自变量之间存在复共线时，最小二乘估计会变“坏”.

为了克服最小二乘估计的上述缺点，相继提出了一些新的估计方法. 岭估计就是其中的一种. 这些估计从形式上看仍然是线性的，但不再保持无偏性，通常称它们为**线性有偏估计**. 它们在理论

上或实用上都是有益的，本节讨论岭回归估计.

在下面的讨论中，我们将以线性回归模型(15.53)～(15.56)为基础. 还假定矩阵 $\boldsymbol{X}$ 已经是中心化，标准化的，并且中心化和标准化了的 $\boldsymbol{X}$ 和 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 具有上述"病态"或复共线特征.

15.5.1 岭回归的定义和性质

岭回归(ridge regresson)是 A E Hoeil 在1962年提出来的，其后由 R W Kennard 系统发展起来的一种改进的最小二乘估计方法.

定义 15.5.1 设 $0\leqslant k<\infty$，称 $\boldsymbol{b}(k)=(b_1(k),b_2(k),\cdots,b_p(k))^T=(\boldsymbol{A}+k\cdot\boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}\triangleq\boldsymbol{W}_k\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$

为 $\boldsymbol{\beta}$ 的**岭回归估计**(ridge regresson). 其中

$$\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\quad \text{是系数矩阵,}$$
$$\boldsymbol{W}_k=(\boldsymbol{A}+k\boldsymbol{I})^{-1}$$

下面说明岭回归估计的性质.

由于 $\boldsymbol{A}$ 的特征根是 $\lambda_1,\lambda_2,\cdots,\lambda_p$，所以 $\boldsymbol{A}+k\boldsymbol{I}$ 的特征根是 $\lambda_1+k,\lambda_2+k,\cdots,\lambda_p+k$. 如果 $\boldsymbol{A}$ 的最小特征根 λ_p 很接近于零，那么 λ_p+k 由于 $k(k\neq 0)$ 是常数接近零的程度就会小些，故有理由希望 $\boldsymbol{b}(k)$ 比 $\boldsymbol{b}$ 性能要好些.

实际上，我们有：

$$\begin{aligned}H(k)&=\boldsymbol{E}(\|\boldsymbol{b}(k)-\boldsymbol{\beta}\|^2)\\&=r_1(k)+r_2(k)\end{aligned}\tag{15.57}$$

其中

$$r_1(k)=\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}$$
$$r_2(k)=k^2\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{A}+k\boldsymbol{I})^{-2}\boldsymbol{\beta}\tag{15.58}$$

此处 $r_1(k)$ 是各回归系数 $\beta_1,\beta_2,\cdots,\beta_p$ 的估计 $b_1(k),b_2(k),\cdots,b_p(k)$ 的方差的平方和，$r_2(k)$ 是 $b(k)$ 的有偏性带来的总偏差.

由式(15.58)可以看出，在 $k\geqslant 0$时，$r_1(k)$是一个下降的函数；且可以证明 $r_2(k)$，在 $k\geqslant 0$时是一个上升的函数，并且 $k=0$ 时，$r_2(k)=0$. 它们的函数图形大致如图15.2示，从图中看出，存在某个 $k>0$，有

$$H(k) < H(0) \tag{15.59}$$

图 15.2

即

$$E(\|\boldsymbol{b}(k)-\boldsymbol{\beta}\|^2) < E(\|\boldsymbol{b}-\boldsymbol{\beta}\|^2) \tag{15.60}$$

也就是说，在某个 k 之下的岭回归估计 $\boldsymbol{b}(k)$的均方差，要比最小二乘估计 $\boldsymbol{b}$ 的均方差来得小.

但是，上述性质只指出存在着这样的 k 值，却并未给出如何计算 k. 实际上，k 的最优值依赖于模型的未知参数 $\boldsymbol{\beta}$ 和 σ^2. 在实际应用时，k 值是不能事先指出，而是必须通过样本来确定的，从这个意义上看，$\boldsymbol{b}(k)$实质上是一个非线性估计.

15.5.2 岭回归估计的计算

如前所述，我们可以用岭回归估计的定义：

$$\boldsymbol{b}(k)=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

来计算 $\boldsymbol{b}(k)$，但是对于不同的 k 值求 $\boldsymbol{b}(k)$，每一次都要计算一个逆矩阵，计算量太大. 下述的计算方法可以减少计算量.

方法15.5.2 计算步骤：

(1) 构造矩阵 $\boldsymbol{U}$

$$\boldsymbol{U}\triangleq\begin{bmatrix}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} & \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}\\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\end{bmatrix}$$

(2) 求特征方程

$$\boldsymbol{A}\boldsymbol{U}=\boldsymbol{\Lambda}\boldsymbol{A}$$

的解 $\boldsymbol{A}$ 与 Λ. 所以 $\boldsymbol{A}$ 的行向量是 $\boldsymbol{U}$ 的特征向量，由此求得 $p+1$阶的正交矩阵 $\boldsymbol{A}$.

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & \cdots & a_{1,p+1} \\ \cdots & \cdots & \cdots \\ a_{p+1,1} & \cdots & a_{p+1,p+1} \end{bmatrix}$$

(3) 给定 k 的值.

(4) 利用下述公式计算岭回归估计

$$b_{j-1}(k)=\frac{-\left(\sum_{i=1}^{p+1} a_{i1}\cdot a_{ij}\Big/(\lambda_i+k)\right)}{\left(\sum_{i=1}^{p+1}\frac{a_{i1}^2}{\lambda_i+k}\right)}$$

$$j=2,3,\cdots,p+1$$

(5) 对不同的 k 值重复(3),(4)步，直到可画出岭迹曲线 $b_{j-1}(k)$来.

(6) 作岭迹分析，求出最优 k_0值，最后确定在最优 k_0值下的岭回归估计 $b(k_0)$.

15.5.3 岭迹分析与最优 k 值的选取

在岭回归中，岭迹分析是了解各自变量的作用及其相互关系的重要途径. 下面用例子说明.

例15.5.3 在图15.3(a)中，$b_i(0)=b_i>0$且值比较大. 从古典回归分析观点看，x_i 对 Y 应是有重要影响的因素. 但 $b_i(k)$的图形，当 k 从0开始增加，显著地下降，迅速地趋向0，因而失去预报的能力，图形呈现出相当的不稳定. 从岭回归的观点看，x_i 对 Y 不起重要作用，甚至可以去掉这个变量.

例15.5.4 在图15.3(b)中，情形正好与情形(a)相反. 开始 $b_i(0)>0$，但值很小. 从古典回归分析看，x_i 对 Y 的作用不大. 随着 k 的增大，$b_i(k)$变为负，从岭回归观点看，x_i 对 Y 有显著影响.

例15.5.5 在图15.3(c)中，从古典回归分析看，x_i 是对 Y 有“正”影响的显著因素；但从岭回归分析看，x_i 对 Y 则是有“负”影响的因素.

例15.5.6 在图15.3(d)中，分别看，系数 $b_1(k)$ 和 $b_2(k)$ 都很不稳定，但其和大体稳定，这表明自变量 x_1 和 x_2 之间存在着相当大的相关性，从变量选择观点看，二者中只保存一个就足够了.

例15.5.7 在图15.3(e)中，岭迹表现紊乱，很不稳定，这时对最小二乘估计的性能应提出疑问.

在15.3(f)图中，与情形 e 相反，最小二乘估计性能良好.

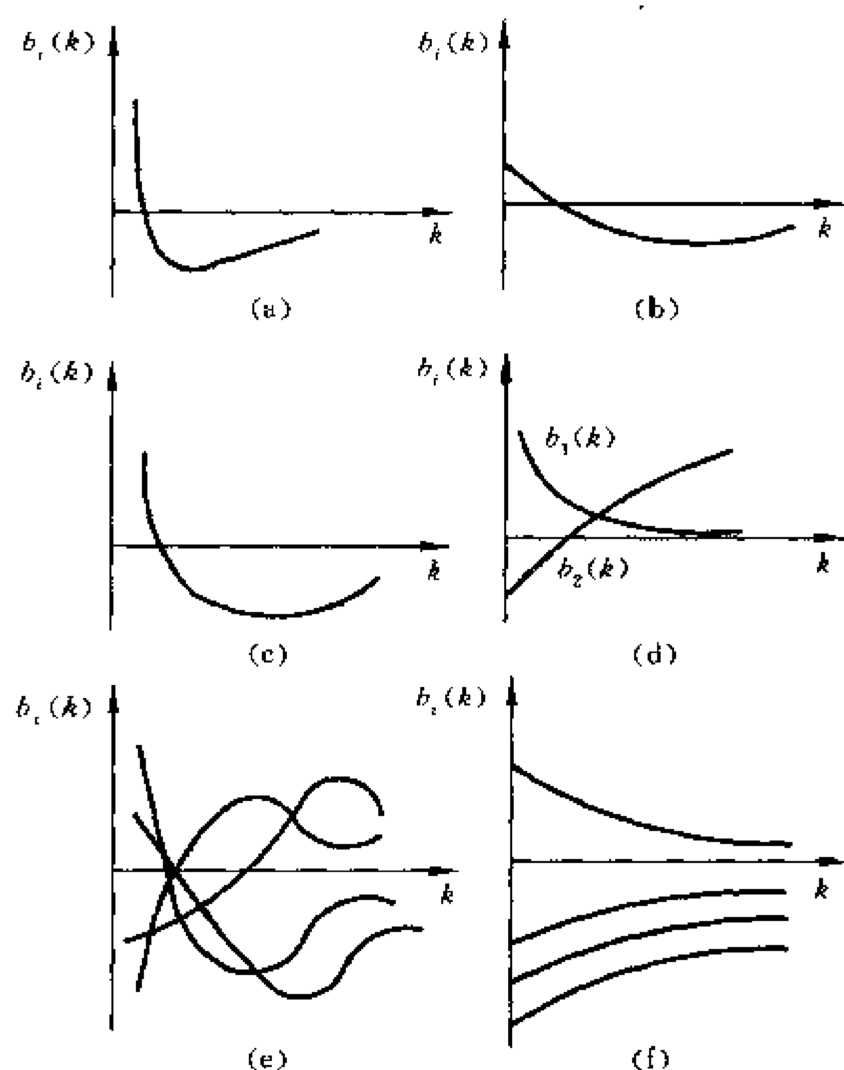


图 15.3

最后，说明最优 k 值的选择方法. 由于最优 k 值和未知模型参数 $\boldsymbol{\beta}$ 和 σ^2 有关，因此在实际应用时往往最终要用样本决定. 还未能从理论上找到一个公认的方法，下面介绍一些有影响的方法.

方法15.5.8 岭迹法

最优 k 值的选择应使各回归系数的估计值都能达到稳定(见图15.3(f)中右端部分)，并且没有不合理的符号，残差平方和也不会增加得太多.

例如在图15.4中，当 k 取 k_0 时，各回归系数都能到达稳定.

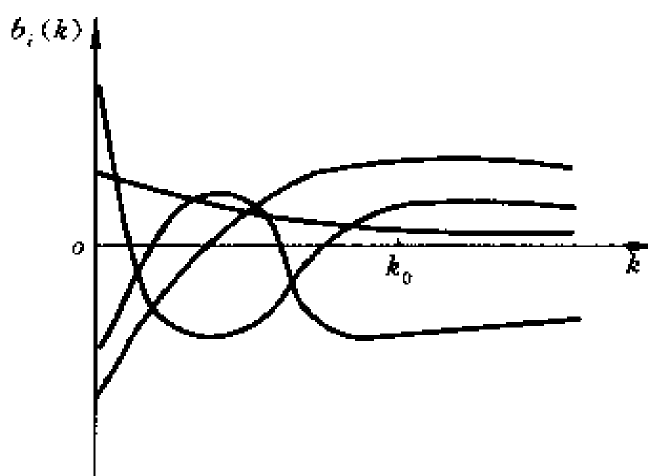


图 15.4

方法15.5.9 选择 k 使均方误差 $H(k)=E(b(k)-\beta)^2$ 最小.

$$H(k)=\sigma^2\sum_{i=1}^{p+1}\frac{\lambda_i}{(\lambda_i+k)^2}+k^2\boldsymbol{\beta}^{\mathrm{T}}[(\boldsymbol{X}^{T}\boldsymbol{X})^{-1}+k\boldsymbol{I}]^{-2}\beta$$

首先用 $\boldsymbol{\beta}$ 和 σ^2 的最小二乘估计 $\boldsymbol{b}$ 和 $\hat{\sigma}^2$ 来代替 $\boldsymbol{\beta}$ 和 σ^2. 对不同的 k 值计算 $H(k)$，描出 $H(k)$ 曲线，选择使 $H(k)$ 达到最小的 k 即为所求.

方法15.5.10 决定 k 值的其它若干方法.

设 $\boldsymbol{X}$ 矩阵已中心化，$\boldsymbol{P}$ 为正交矩阵使

$$\boldsymbol{P}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{P}^{\mathrm{T}}=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_{p+1})$$

称 $\boldsymbol{\alpha}=\boldsymbol{P\beta}$ 为典则参数，记 $\boldsymbol{Z}=\boldsymbol{XP}^{\mathrm{T}}$，则原模型变形为

$$Y = Z\alpha + \varepsilon$$

α 的最小二乘估计和岭估计分别为

$$\hat{\alpha} = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}Y = \Lambda^{-1}PXX^{\mathrm{T}}Y.$$

$$\hat{\alpha}(k) = ((Z^{\mathrm{T}}Z) + kI)^{-1}PX^{\mathrm{T}}Y = (\Lambda + kI)^{-1}PX^{\mathrm{T}}Y.$$

下面的四种求 k 值的公式都是通过 $\hat{\alpha}$ 表示的.

$$k = \hat{\sigma}^2/\max \hat{\alpha}_i^2. \tag{15.61}$$

$$k = (p+1)\hat{\sigma}^2 \Big/ \sum_{i=1}^{p+1} \lambda_i \hat{\alpha}_i^2. \tag{15.62}$$

$$k = (p+1)\hat{\sigma}^2 \Big/ \sum_{i=1}^{p+1} \hat{\alpha}_i^2. \tag{15.63}$$

当 $Q = \|\hat{\alpha}\|^2 - \hat{\sigma}^2 \sum_{i=1}^{p+1} \frac{1}{\lambda_i} > 0$ 时，选 k，使 $\|\hat{\alpha}(k)\|^2 \approx Q$，

$$\tag{15.64}$$

当 $Q = \|\hat{\alpha}\|^2 - \hat{\sigma}^2 \sum_{i=1}^{p+1} \frac{1}{\lambda_i} < 0$ 时，选 $k = 0$.

例15.5.11 记 $S = X^{\mathrm{T}}X$，由于 X 已经中心化和标准化，故 S 为相关矩阵，其特征根为

3.692，1.542，1.293，1.046，0.972，0.659，0.357，0.220，0.152，0.068.

最后一个特征根为0.068，很接近于0. 通常最小二乘估计不会给出良好结果. 现用岭回归分析，对15个 k 值算出 $b_i(k)$，描出岭迹. 如图15.5和图15.6所示.

从图15.5看，最小二乘估计稳定性很差，这反映在当 k 与0略有偏差时，$b_i(k)$ 与 $b_i(0)$ 就有较大差距，特别是 $|b_5|$ 与 $|b_6|$ 下降最多. 因素 x_5 的回归系数的最小二乘估计 $b_5 = b_5(0)$ 为负回归系数中绝对值最大的. 但当 k 增大时，$b_5(k)$ 迅速上升且变为正的，与此相反，对因素 x_6，b_6 为正且绝对值最大. 但当 k 增大时，$b_6(k)$ 迅速下降. 再考虑到 x_6，x_5 的相关系数达到0.84，因此这两个因素可近似

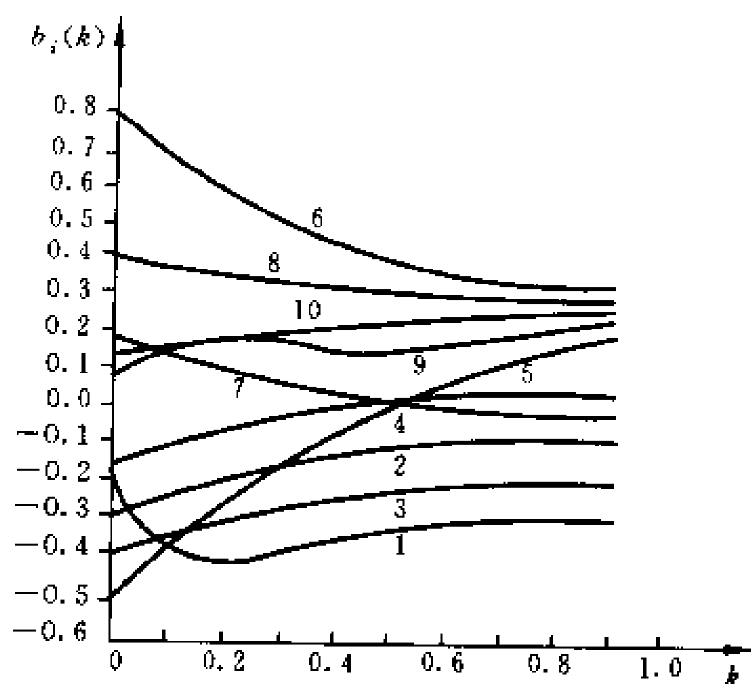


图 15.5

岭迹(10个因素)

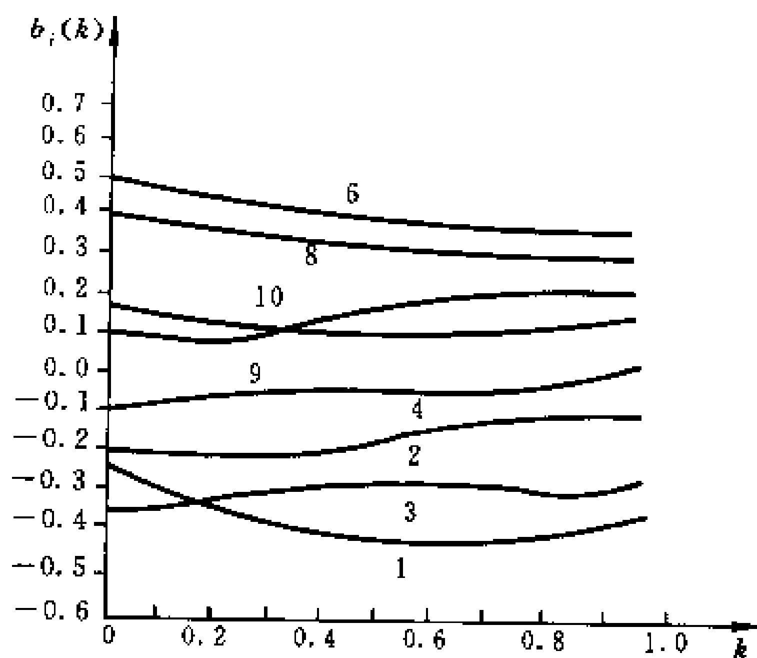


图 15.6

岭迹(去掉因素 x_5, x_7)

地合并成一个因素，再看 x_7，它的回归系数估计值 $b_7=b_7(0)$ 绝对值偏高；当 k 增大时，$b_7(k)$ 很快接近于0，这意味着 x_7 实际上对 Y 无多大作用. 从整体上看，当 k 达到0.2—0.3的范围时，各个 $b_i(k)$ 已大体上趋于稳定. 因此在这区间上取一个 k 值做岭回归. 从岭回归观点看，本例中 $b_5(k)$ 和 $b_7(k)$ 当 k 从0增加时，很快趋于0，于是它们很自然地应被舍去. 去掉 x_5，x_7 后. 重作岭回归分析，结果如图(15.6)所示. 基本稳定(即图形波动不大)，所以去掉 x_5 和 x_7 是合理的.

16 方差分析

16.1 引言

方差分析(analysis of variance)是分析试验数据的一种方法.对于抽样测得的实验数据,由于观测条件不同(即同一因素的不同水平或不同因素的各个水平)会引起试验结果有所不同;另一方面由于各种随机因素的干扰,实验结果也会有所不同,由观测条件不同所引起的实验结果的差异是系统性的,而随机因素引起的差异是偶然性的.在一个试验中,各种**因素**或称**因子**(factors)综合在一起,它们互相制约,互相依存.方差分析的目的在于从实验数据中分析出各个因素的影响及各个因素间的交互影响,以确定各个因素作用的大小,从而把由于观测条件不同而引起实验结果的不同与由于随机因素而引起实验结果的差异用数量的形式区别开来,以确定在试验中有没有系统性的因素(即不同的因素或同一因素的各个不同水平)在起作用.

方差分析在化工、冶金、农业、地质、石油勘探,以及卫星图象分析等许多领域中都获得了应用.本章将着重介绍单因子和双因子方差分析.

16.2 单因子方差分析

首先讨论单因子方差分析(one-way analysis of variance)的情况.在一个试验中,影响试验的因素往往很多,但是如果其它因素能够控制在一定范围之内,为了研究方便,可以认为这时只有一

个因素在变化，仅仅分析这一个因素对试验的影响，这就是单因子方差分析.

在试验中，为了评价试验的性质需要进行多次测量，我们把测量的结果称为性能指标，或简称为指标. 在试验中影响指标的因素称为**因子**(facfor)，因子所处的状态，所取的等级称为因子水平. 例如，我们要研究某化工产品的工艺条件，考虑温度对试验的影响，这时温度就是因子，如果让温度取值为 50°, 60°, 70°, 80°，那么这些温度等级就是**因子水平**(factor level).

16.2.1 试验次数相等的方差分析

假定因子 A 有 m 个水平，分别记为 $A_1, A_2, \cdots, A_m$. 在每一种水平下，做 k 次试验，在每次试验后可得一试验值，记做 x_{ij}，它表示在第 i 个水平下的第 j 个试验值($i=1,2,\cdots,m; j=1,2,\cdots,k$). 结果用表 16.1 表示.

表 16.1

因子水平 \ 试验次数	1	2	$\cdots$	j	$\cdots$	k	$\sum\limits_{j=1}^{k} x_{ij}$	$\overline{x_i} = \frac{1}{k}\sum\limits_{j=1}^{k} x_{ij}$
A_1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1k}	$\sum x_{1j}$	$\overline{x_1}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2k}	$\sum x_{2j}$	$\overline{x_2}$
$\vdots$	$\vdots$						$\vdots$	$\vdots$
A_i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{ik}	$\sum x_{ij}$	$\overline{x_i}$
$\vdots$	$\vdots$						$\vdots$	$\vdots$
A_m	x_{m1}	x_{m2}	$\cdots$	x_{mj}	$\cdots$	x_{mk}	$\sum x_{mj}$	$\overline{x_m}$

为了考察因子 A 对试验结果是否有显著的影响，我们把因子 A 的 m 个水平 $A_1, A_2, \cdots, A_m$ 看成是 m 个正态总体，而 $x_{ij}, j=1,$

$\cdots k$ 为取自第 i 个总体的第 j 个样本，$i=1,\cdots,m$. 因此可设 $X_{ij}\sim N(a_i,\sigma^2)$，$i=1,2,\cdots,m$，$j=1,2,\cdots,k$.

可以认为 $a_i=\mu+\varepsilon_i$，ε_i 是因子 A 的第 i 个水平 A_i 所引起的差异. 因此检验因子 A 的各水平之间是否有显著的差异，相当于做下列检验：

$$H_{01}:\ a_1=a_2=\cdots=a_m=\mu, \tag{16.1}$$

或者相当于做检验：$H_{01}:\ \varepsilon_1=\varepsilon_2=\cdots=\varepsilon_m=0$.

如何把由于因子水平的不同而造成结果的差异与由于随机因素的影响而造成结果的差异两者从数量上区分开来就要用到下面的平方和分解公式.

公式 16.2.1 平方和分解公式(decomposition formalae of sum of square)总的离差平方和 S_T(total sum of squares)可以进行分解：

$$S_T=S_e+S_A,$$

其中

$$S_T=\sum_{i=1}^{m}\sum_{j=1}^{k}(x_{ij}-\overline{x})^2,$$

$$S_A=\sum_{i=1}^{m}\sum_{j=1}^{k}(\overline{x_i}-\overline{x})^2=k\sum_{i=1}^{m}(\overline{x_i}-\overline{x})^2,$$

$$S_e=\sum_{i=1}^{m}\sum_{j=1}^{k}(x_{ij}-\overline{x_i})^2,$$

$$\overline{x_i}=\frac{1}{k}\sum_{j=1}^{k}x_{ij},$$

$$\overline{x}=\frac{1}{m}\sum_{i=1}^{m}\overline{x_i}=\frac{1}{mk}\sum_{i=1}^{m}\sum_{j=1}^{k}x_{ij}.$$

S_T 称为总离差平方和，它是所有观察值 x_{ij} 与其总平均值 $\overline{x}$ 之差的平方和，是描述全部数据离散程度的数量指标. 由前述 x_{ij} 是服从正态分布的随机量，当(16.1)H_{01}成立时，x_{ij}是独立同分

布，同正态分布的随机量，可以证明$\frac{S_T}{\sigma^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2}{\sigma^2}$是服从自由度$f_T = mk - 1$的$\chi^2$分布.

$S_e = \sum_{i=1}^{m}\sum_{j=1}^{k}(x_{ij}-\bar{x}_i)^2$是观察值与组内平均值$\bar{x}_i$之差的平方和，称为组内平方和或误差平方和. 它反映了组内（在同一水平之下）样本的随机波动. 注意到S_e是非负的二次型，并且在mk个平方和中有m个约束条件$\sum_{j=1}^{k}(x_{ij}-\bar{x}_i) = 0, i = 1,2,\cdots,m$，所以$\sum_{i=1}^{m}\sum_{j=1}^{k}(x_{ij}-\bar{x}_i)^2$的自由度$f_e = mk - m$.

$S_A = k\sum_{i=1}^{m}(\bar{x}_i - \bar{x})^2$是组内平均值$\bar{x}_i$与总平均值之差的平方和，称为组间平方和. 它在一定程度上反映了因子各个水平不同而引起的差异. 注意到S_A是非负的二次型，有m个平方和相加，而且有一个约束条件：$\sum_{i=1}^{m}(\bar{x}_i - \bar{x}) = 0$，所以$S_A$的自由度$f_A = m-1$.

定理 16.2.2 平方和分解定理

设Q服从 自由度为n的χ^2分布，又

$$Q_1 + Q_2 + \cdots + Q_k = Q$$

其中$Q_i(i=1,2,\cdots,k)$是秩为$f_i(i=1,2,\cdots,k)$的非负二次型，则$Q_i, i=1,2,\cdots,k$相互独立，且分别服从自由度为$f_i(i=1,2,\cdots,k)$的χ^2分布的充要条件是

$$f_1 + f_2 + \cdots + f_k = n$$

注意到平方和分解公式 16.2.1 在(16.1)H_{01}为真的假定下$S_T/\sigma \sim \chi^2(mk-1)$，$S_e/\sigma$，$S_A/\sigma^2$是非负二次型，其秩分别为$mk-m$，$m-1$. 由于$(mk-m)+m-1=mk-1$，即$f_T = f_e + f_A$，所以根据平方和分解定理 16.2.2 知$S_e/\sigma^2$与$S_A/\sigma^2$相互独立，且分别服

从自由度为 f_e 及 f_A 的 χ^2 分布.

平方和分解公式 $S_T=S_e+S_A$ 向我们揭示了这样的事实,即观察值关于其总平均值之间的差异可以看成是由两部分构成的,一部分是组内平方和,或称为误差平方和,它反映了因随机因素的作用引起的差异.另一部分是组间平方和,它是由因子的各个水平不同而引起的差异.因此 S_A 与 S_e 的比值就反映了两种差异所占的比重,若 S_A 与 S_e 之比值越大,说明因子的各个水平不同引起的差异显著.因此,统计量

$$F=\frac{S_A/m-1}{S_e/m(k-1)} \tag{16.2}$$

可用来检验因子的效应是否显著.

在假设 H_{01}(16.1)成立时,量 $S_T/\sigma^2,S_e/\sigma^2,S_A/\sigma^2$ 都是 χ^2 变量,其自由度分别是 $f_T=mk-1,f_e=mk-m,f_A=m-1$.并且成立关系:

$$f_T=f_e+f_A \tag{16.3}$$

根据定理 16.2.2:在假设 H_{01}成立时,S_e 与 S_A 是相互独立的,其自由度分别为 f_e 与 f_A 的 χ^2 变量.从而,统计量

$$F\sim F(m-1,m(k-1)) \tag{16.4}$$

现在可以给出检验假设 H_{01}的规则如下:对于给定的显著性水平 α,由 F 分布表查出自由度为$(m-1,m(k-1))$的临界值 F_α,如果 $F>F_\alpha$,则拒绝原假设 H_{01},此时说明因子对指标起显著影响;如果 $F\leqslant F_\alpha$,则接受 H_{01},此时说明因子 A 的不同水平对试验结果影响不显著.

方法 16.2.3　计算步骤

(1) 为了减少计算错误,常常先对原始观测数据做下述线性变换:

$$y_{ij}=\frac{x_{ij}-c}{d},\quad (i=1,2,\cdots,m;j=1,2,\cdots,k)$$

式中的 c,d 是常数.

(2) 将新数据列成如表 16.2 的计算表格

表 16.2

水平 \ 次数	1 2… k	和 $\sum$	和平方 $(\sum)^2$	平方和 $\sum^2$
A_1	$y_{11}\ y_{12}\cdots\ y_{1k}$	$\sum_{j=1}^{k} y_{1j}$	$\left(\sum_{j=1}^{k} y_{1j}\right)^2$	$\sum_{j=1}^{k} y_{1j}^2$
A_2	$y_{21}\ y_{22}\cdots\ y_{2k}$	$\sum_{j=1}^{k} y_{2j}$	$\left(\sum_{j=1}^{k} y_{2j}\right)^2$	$\sum_{j=1}^{k} y_{2j}^2$
⋮	⋮		⋮	⋮
A_m	$y_{m1}\ y_{m2}\cdots\ y_{mk}$	$\sum_{j=1}^{k} y_{mj}$	$\left(\sum_{j=1}^{k} y_{mj}\right)^2$	$\sum_{j=1}^{k} y_{mj}^2$
总和		$\sum_{i=1}^{m}\left(\sum_{j=1}^{k} y_{ij}\right)$	$\sum_{i=1}^{m}\left(\sum_{j=1}^{k} y_{ij}\right)^2$	$\sum_{i=1}^{m}\left(\sum_{j=1}^{k} y_{ij}^2\right)$

(3) 按下述公式进行计算

$$P=\frac{1}{mk}\left(\sum_{i=1}^{m}\sum_{j=1}^{k} y_{ij}\right)^2$$

$$Q=\frac{1}{k}\sum_{i=1}^{m}\left(\sum_{j=1}^{k} y_{ij}\right)^2$$

$$R=\sum_{i=1}^{m}\sum_{j=1}^{k} y_{ij}^2$$

$$S'_T=R-P, \qquad f_T=mk-1$$

$$S'_A=Q-P, \qquad f_A=m-1$$

$$S'_e=R-Q, \qquad f_e=m(k-1)$$

$$F=\frac{S'_A/f_A}{S'_e/f_e}$$

列出方差分析表.

表 16.3　单因子方差分析表

方差来源	平方和 S'	自由度 f	平均平方和（均方）	F 值	显著性
组间	S'_A	f_A	S'_A/f_A	$\dfrac{S'_A/f_A}{S'_e/f_e}$	
组内	S'_e	f_e			
总和	S'_T	f_T			

（4）进行统计检验

对于给定的显著性水平 α，由 F 分布表及自由度 (f_A, f_e) 查出相应的临界值 F_α，如果 $F > F_\alpha$，则拒绝假设 H_{01}，认为 A 因子作用显著；如果 $F \leqslant F_\alpha$，则接受假设 H_{01}，认为因子 A 的作用不显著.

例 16.2.4　要考察染整工艺对布的缩水率是否有显著的影响. 试验中采取了 5 种染整工艺，每种工艺处理了 4 块布样，测得缩水率的百分数如表 16.4 所示.

表　16.4

染整工艺 \ 布样号	1	2	3	4
A_1	4.3	7.8	3.2	6.5
A_2	6.1	7.3	4.2	4.1
A_3	6.5	8.3	8.6	8.2
A_4	9.3	8.7	7.2	10.1
A_5	9.5	8.8	11.4	7.8

（1）做线性变换

$$y_{ij} = x_{ij} - 5, \qquad \begin{pmatrix} i = 1,2,\cdots,5 \\ j = 1,2,\cdots,4 \end{pmatrix}$$

（2）将新数据列成表 16.5 的计算格式

按表 16.4 计算出表 16.5.

表 16.5

染整工艺 \ 布样号	1	2	3	4	$\sum$	$(\sum)^2$	$\sum^2$
A_1	−0.7	2.8	−1.8	1.5	1.8	3.24	13.82
A_2	1.1	2.3	−0.8	−0.9	1.7	2.89	7.95
A_3	1.5	3.3	3.6	3.2	11.6	134.56	36.34
A_4	4.3	3.7	2.2	5.1	15.3	234.09	63.03
A_5	4.5	3.8	6.4	2.8	17.5	306.25	63.49
总和					47.9	681.03	204.63

按方法 16.2.3 进行计算

$$P=\frac{1}{5\times4}(47.9)^2=114.72$$

$$Q=\frac{1}{4}681.03=170.26$$

$$R=204.63$$

$$S_T'=R-P=214.63-114.72=99.91$$

$$S_A'=Q-P=170.26-114.72=55.54$$

$$S_e'=R-Q=204.63-170.26=34.37$$

$$f_T=5\times4-1=19$$

$$f_A=5-1=4$$

$$f_e=5(4-1)=15$$

$$F=\frac{55.54/4}{34.37/15}=6.06$$

（3）列方差分析表如表 16.6.

表 16.6

方差来源	平方和	自由度	均方	F	显著值
组间	55.54	4	13.885	6.06	* *
组内	34.37	15	2.291		
总和	99.91				

(4) 统计检验

由自由度($f_A=4, f_e=15$),显著性水平 $\alpha=0.05$,查 F 分布表得 $F_{0.05}=3.1$,由于计算值 6.1>3.1 说明不同染整工艺的差别显著,且第一、二两种染整工艺(即 A_1, A_2)对布料缩水率的影响最小.

16.2.2 试验次数不等的方差分析

试验中,有时获得的数据不那么整齐,各个水平下的试验次数可能不等,此时计算公式要作相应的调整.

方法 16.2.5 计算步骤

(1) 列出统计分析数据表如表 16.7

表 16.7

次数 / 水平	1 2 … k_1 … k_2 … k_j … k_n	$\sum$	$\frac{1}{k_i}(\sum)^2$	$\sum^2$
A_1	x_{11} x_{12} … x_{1k_1}			
A_2	x_{21} x_{22} … … x_{2k_2}			
⋮				
A_i	x_{i1} x_{i2} … x_{ik_j}			
⋮	…			
A_m	x_{m1} x_{m2} … x_{mk_n}			

(2) 对表中的每一行求和 $\sum$，和的平方的平均值 $\frac{1}{k_i}(\sum)^2$，以及平方和 $\sum^2$.

(3) 按下述公式进行计算

$$n=\sum_{i=1}^{m}k_i$$

$$P=\frac{1}{n}\left(\sum_{i=1}^{m}\sum_{j=1}^{k_i}x_{ij}\right)^2$$

$$Q=\sum_{i=1}^{m}\frac{1}{k_i}\left(\sum_{j=1}^{k_i}x_{ij}\right)^2$$

$$R=\sum_{i=1}^{m}\sum_{j=1}^{k_i}x_{ij}^2$$

$$S_e=R-Q$$

$$S_A=Q-P$$

$$S_{总}=R-P$$

$$F=\frac{S_A/(m-1)}{S_e/(n-m)}$$

(4) 统计检验

根据显著性水平 α，以及自由度 $(m-1,n-m)$ 查 F 分布表，得临界值 F_α，若计算值 $F>F_\alpha$，则认为 A 的作用显著；如果 $F\leqslant F_\alpha$ 则认为 A 的作用不显著.

16.3 双因子方差分析

在实际问题中，影响试验结果的因素往往不只一个. 这时需要用双因子乃至多因子的方差分析，从而确定在众多的因素中哪些因素是主要的，它们对试验结果的影响是否显著，以及它们之间是否有交互作用. 下面，我们将详细说明双因子方差分析(two way

analysis of variance)的基本思想和计算方法.

16.3.1 无交互作用的双因子方差分析

假定 A,B 是影响试验结果的两个因子,因子 A 有 n 个水平:$A_1,A_2,\cdots,A_n$;因子 B 有 m 个水平:$B_1,B_2,\cdots,B_m$. 在 A 的 n 个水平与 B 的 m 个水平的每种组合下作一次试验,可得数据表 16.8.

表 16.8

	B_1	B_2	$\cdots$	B_m
A_1	x_{11}	x_{12}	$\cdots$	x_{1m}
A_2	x_{21}	x_{22}	$\cdots$	x_{2m}
$\vdots$	$\vdots$			
A_n	x_{n1}	x_{n2}	$\cdots$	x_{nm}

现在,我们假定在每种组合下的试验结果 x_{ij}服从均值为 a_{ij}、方差为 σ^2 的正态分布,即 $x_{ij}\sim N(a_{ij},\sigma^2)$,将 a_{ij}写成

$$a_{ij} = \mu + \alpha_i + \beta_j, \quad \left.\begin{array}{l} i = 1,2,\cdots,n \\ j = 1,2,\cdots,m \end{array}\right\} \tag{16.5}$$

其中 α_i 表示因子 A 的第 i 个水平的影响,β_j 表示因子 B 的第 j 个水平的影响. 这个模型的含义是说,A,B 因子的不同水平对均值的影响是累加效果(指 $\alpha_i+\beta_j$). 因此这种模型又称为可加效应模型. 这时判定因子 A 和因子 B 对试验指标影响是否显著的问题,相当于做下列检验 H_0.

$$H_0:\ H_{01} \cap H_{02} \tag{16.6}$$

其中 $H_{01}:\ \alpha_1=\alpha_2=\cdots=\alpha_n=0$

$H_{02}:\ \beta_1=\beta_2=\cdots=\beta_m=0$

H_0 也等价于 $a_{11}=a_{12}=\cdots=a_{ij}=a_{nm}$

和单因子情况类似,能够证明下列公式成立

公式 16.3.1　分解公式

$S_T=S_A+S_B+S_e$,

其中

$$S_T=\sum_{j=1}^{m}\sum_{i=1}^{n}(x_{ij}-\overline{x})^2,$$

$$S_A=\sum_{i=1}^{n}\sum_{j=1}^{m}(\overline{x_{i\cdot}}-\overline{x})^2,$$

$$S_B=\sum_{i=1}^{n}\sum_{j=1}^{m}(\overline{x_{\cdot j}}-\overline{x})^2,$$

$$S_e=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}-\overline{x_{i\cdot}}-\overline{x_{\cdot j}}+\overline{x})^2,$$

$$\overline{x_{i\cdot}}=\frac{1}{m}\sum_{j=1}^{m}x_{ij},$$

$$\overline{x_{\cdot j}}=\frac{1}{n}\sum_{i=1}^{n}x_{ij},$$

$$\overline{x}=\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}.$$

当 $H_{01}\cap H_{02}$ 成立时,即 $x_{ij}\sim N(\mu,\sigma^2)$.

可知　$S_T/\sigma^2=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}-\overline{x})^2/\sigma^2\sim\chi^2(mn-1)$.

$S_A=\sum_{i=1}^{n}\sum_{j=1}^{m}(\overline{x_{i\cdot}}-\overline{x})^2=m\sum_{i=1}^{n}(\overline{x_{i\cdot}}-\overline{x})^2$,是非负二次型,其中有一个约束条件:$\sum_{i=1}^{n}(\overline{x_{i\cdot}}-\overline{x})=0$,所以 S_A 的秩为 $f_A=n-1$.

$S_B=\sum_{i=1}^{n}\sum_{j=1}^{m}(\overline{x_{\cdot j}}-\overline{x})^2=n\sum_{j=1}^{m}(\overline{x_{\cdot j}}-\overline{x})^2$ 是非负二次型,有一个约束条件 $\sum_{j=1}^{m}(\overline{x_{\cdot j}}-\overline{x})=0$,所以 S_B 的秩为 $f_B=m-1$.

$S_e = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \overline{x_{i\cdot}} - \overline{x_{\cdot j}} + \overline{x})^2$，有约束条件：$\sum_{j=1}^{m}(x_{ij} - \overline{x_{i\cdot}} - \overline{x_{\cdot j}} + \overline{x}) = 0, i = 1,2,\cdots,n$，共 n 个，$\sum_{i=1}^{n}(x_{ij} - \overline{x_{i\cdot}} - \overline{x_{\cdot j}} + \overline{x}) = 0, j = 1,2,\cdots,m$，共 m 个，它们中都包含 $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \overline{x_{i\cdot}} - \overline{x_{\cdot j}} + \overline{x}) = 0$，即约束条件有 $n+m-1$ 个，所以 S_e 的秩为 $mn-(m+n-1)=(m-1)(n-1)$.

综上所述 S_T,S_A,S_B,S_e 的秩分别为：

$f_T=mn-1, f_A=n-1, f_B=m-1, f_e=(m-1)(n-1)$.

由于 $S_T/\sigma^2=S_A/\sigma^2+S_B/\sigma^2+S_e/\sigma^2$，而且 $f_T=f_A+f_B+f_e$.

根据平方和分解定理，在 $H_{01}\cap H_{02}$成立的假定下，S_A,S_B,S_e 相互独立，且 $S_A/\sigma^2\sim\chi^2(n-1)$；$S_B/\sigma^2\sim\chi^2(m-1)$；$S_e/\sigma^2\sim\chi^2(m-1)(n-1)$.

公式 16.3.1 告诉我们，总的离差平方和可以分解为三个部分：S_A 反映了因子 A 的各水平之间的差异；S_B 反映了因子 B 的各水平之间的差异；S_e 反映了在交互作用不显著时试验本身误差的大小. 因此可利用 S_A 与 S_e 的比值大小来考查因子 A 的各水平差异是否显著，利用 S_B 与 S_e 之比值的大小来考查因子 B 的各水平差异是否显著. 下面的统计量就是用来做显著性检验的统计量.

$$F_A = \frac{S_A/(n-1)}{S_e/(n-1)(m-1)} \sim F(n-1,(n-1)(m-1)) \tag{16.7}$$

$$F_B = \frac{S_B/(m-1)}{S_e/(n-1)(m-1)} \sim F(m-1,(n-1)(m-1)) \tag{16.8}$$

根据显著性水平 α，自由度(f_A,f_e)，(f_B,f_e)，由 F 分布表，分别查出统计量 F_A 和 F_B 的临界值，再分别检验因子 A 与因子 B 对试验是否有显著影响.

公式 16.3.2　计算公式和计算步骤

(1) 为了便于计算,经常对原始数据作下述的线性变换:

$$y_{ij}=\frac{x_{ij}-c}{d}\qquad(i=1,\cdots,n;j=1,\cdots,m)$$

式中 c,d 均为常数.

(2) 列表,并分别对行和列求出和 $\sum$,和平方 $\left(\sum\right)^2$,以及平方和 $\sum^2$. 如表 16.9 所示.

表　16.9

A的水平 \ B的水平	B_1	…	B_m	和 $\sum$	和平方 $\left(\sum\right)^2$	平方和 $\sum^2$
A_1	y_{11}	…	y_{1m}	$\sum_{j=1}^{m}y_{1j}$	$\left(\sum_{j=1}^{m}y_{1j}\right)^2$	$\sum_{j=1}^{m}y_{1j}^2$
⋮	⋮	…	⋮	⋮	⋮	⋮
A_n	y_{n1}	…	y_{nm}	$\sum_{j=1}^{m}y_{nj}$	$\left(\sum_{j=1}^{m}y_{nj}\right)^2$	$\sum_{j=1}^{m}y_{nj}^2$
和 $\sum$	$\sum_{i=1}^{n}y_{i1}$	…	$\sum_{i=1}^{n}y_{im}$	$\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}$	$\sum_{i=1}^{n}\left(\sum_{j=1}^{m}y_{ij}\right)^2$	$\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}^2$
和平方 $\left(\sum\right)^2$	$\left(\sum_{i=1}^{n}y_{i1}\right)^2$	…	$\left(\sum_{i=1}^{n}y_{im}\right)^2$	$\sum_{j=1}^{m}\left(\sum_{i=1}^{n}y_{ij}\right)^2$		
平方和 $\sum^2$	$\sum_{i=1}^{n}y_{i1}^2$	…	$\sum_{i=1}^{n}y_{im}^2$	$\sum_{j=1}^{m}\sum_{i=1}^{n}y_{ij}^2$		

(3) 按下列公式进行计算

$$P=\frac{1}{nm}\left(\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}\right)^2,$$

$$Q=\frac{1}{m}\sum_{i=1}^{n}\left(\sum_{j=1}^{m}y_{ij}\right)^2,$$

$$R=\frac{1}{n}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}y_{ij}\right)^2,$$

$$T=\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}^2,$$

$$S'_{总}=T-P, \qquad f_{总}=n\cdot m-1,$$

$$S'_A=Q-P, \qquad f_A=n-1,$$

$$S'_B=R-P, \qquad f_B=m-1,$$

$$S'_e=T-Q-R+P, \qquad f_e=(n-1)(m-1),$$

$$F_A=\frac{S'_A/f_A}{S'_e/f_e},F_B=\frac{S'_B/f_B}{S'_e/f_e}.$$

（4）列出方差分析表如表 16.10.

表 16.10　无重复双因子方差分析表

方差来源	平方和 S	自由度 f	（均方）平均平方和 $\overline{S}$	F 值	显著性
A 因子的作用	S'_A	f_A	S'_A/f_A	F_A	
B 因子的作用	S'_B	f_B	S'_B/f_B	F_B	
误差	S'_e	f_e	S'_e/f_e		
总和	$S'_{总}$	$f_{总}$			

（5）进行统计检验.

例 16.3.3　影响某试验的指标有两个因子 A 与 B，各取 4 个水平，其试验数据如表 16.11 所示，现在问，哪一种因子对试验的影响显著.

表 16.11

因子 A \ 因子 B	B_1	B_2	B_3	B_4
A_1	2.20	3.76	3.72	4.17
A_2	2.18	4.72	4.85	5.50
A_3	3.00	4.95	4.20	5.50
A_4	5.10	7.10	8.30	6.31

(1) 对试验数据作如下的变换，令

$$y_{ij} = x_{ij} - 4.$$

(2) 列表计算如表 16.12.

表 16.12

因子 A \ 因子 B	B_1	B_2	B_3	B_4	$\sum$	$(\sum)^2$
A_1	−1.80	−0.24	−0.28	0.17	−2.15	4.62
A_2	−1.82	0.72	0.85	1.50	1.25	1.56
A_3	−1	0.95	0.20	1.50	1.65	2.72
A_4	1.10	3.10	4.30	2.31	10.81	116.86
$\sum$	−3.52	4.53	5.07	5.48	11.56	125.76
$(\sum)^2$	12.39	20.52	25.70	30.03	88.64	
$\sum^2$	8.76	11.09	19.33	9.87	49.05	

(3) 按公式 16.3.2 进行计算

$$P = \frac{1}{4 \times 4}(11.56)^2 = 8.35$$

$$Q = \frac{1}{4}(125.76) = 31.44$$

$$R = \frac{1}{4}(88.64) = 22.16$$

$$T = 49.05$$

$$S'_{总} = T - P = 49.05 - 8.35 = 40.70$$

$$S'_A = Q - P = 31.44 - 8.35 = 23.09$$

$$S'_B = R - P = 22.16 - 8.35 = 13.81$$

$$S'_e = T - Q - R + P = 3.80$$

$$f_{总} = 16 - 1 = 15$$

$$f_A = 4 - 1 = 3$$

$$f_B = 4 - 1 = 3$$

$$f_e = 3 \times 3 = 9$$

(4) 方差分析表如表 16.13.

表 16.13

方差来源	平方和 S	自由度 f	平均离差平方和 $\overline{S}$	F 值	显著性
因子 A	23.09	3	7.70	$F_A=18.33$	显著
因子 B	13.81	3	4.60	$F_B=10.95$	显著
误差	3.80	9	0.42		
总和	40.70	15			

(5) 统计检验

由显著性水平 $\alpha=0.01$ 及自由度(3,9)查 F 分布表得临界值 $F_{0.01}=6.99$.

计算值 $F_A=18.33>6.99$,故因子 A 对试验影响显著;又 $F_B=10.95>6.99$,故因子 B 对试验影响也显著.

16.3.2 有交互作用的双因子方差分析

设有因子 A 和因子 B,因子 A 取 n 个水平,因子 B 取 m 个水

平. 在 A 的 n 个水平和 B 的 m 个水平的各种组合下,各进行了 r 次试验,x_{ijk}表示试验在 A_iB_j 条件下作第 k 次重复试验的数据. 现在有表 16.14 所示数据表.

表 16.14

A 的水平 \ B 的水平	B_1	B_2	…	B_m
A_1	x_{111} x_{112} ⋮ x_{11r}	x_{121} x_{122} ⋮ x_{12r}	…	x_{1m1} x_{1m2} ⋮ x_{1mr}
A_2	x_{211} x_{212} ⋮ x_{21r}	x_{221} x_{222} ⋮ x_{22r}	…	x_{2m1} x_{2m2} ⋮ x_{2mr}
⋮	⋮	⋮	⋮	⋮
A_n	x_{n11} x_{n12} ⋮ x_{n1r}	x_{n21} x_{n22} ⋮ x_{n2r}	…	x_{nm1} x_{nm2} ⋮ x_{nmr}

假定 x_{ijk}(表示在因子 A 的第 i 个水平,因子 B 的第 j 个水平下做第 k 次试验所得数据)服从 $N(a_{ijk},\sigma^2)$. 而 a_{ijk}可以写成:

$$a_{ijk} = \mu + \alpha_i + \beta_j + r_{ij} \tag{16.9}$$

其中 α_i 表示因子 A 第 i 个水平的单独作用引起的差异,β_j 表示因子 B 的第 j 个水平的单独作用引起的差异,r_{ij}表示因子 A 第 i 个水平与因子 B 的第 j 个水平的交互作用引起的差异,因此检验因子 A、因子 B 及其交互作用的影响的问题等价于做下列检验

$$H_0:\quad H_{01} \cap H_{02} \cap H_{03} \tag{16.10}$$

其中 H_{01},H_{02},H_{03}分别为下列检验:

$$H_{01}: \alpha_1=\alpha_2=\cdots=\alpha_n=0$$

$$H_{02}: \beta_1=\beta_2=\cdots=\beta_m=0$$

$$H_{03}: r_{11}=r_{12}=\cdots=r_{ij}=\cdots=r_{nm}=0$$

等价于 $H_0: a_{111}=a_{112}=\cdots=a_{ijk}=\cdots=a_{n,m,r}$

有交互作用的双因子方差分析模型的分析思想和前一节类似，将总的离差平方和分解成 4 个部分：(1) 因子 A 单独作用的结果，(2) 因子 B 单独作用的结果，(3) A 与 B 的交互作用的结果，记作 $A\times B$，(4) 试验误差. 用平方和分解公式表示，则有

$$S_T = S_A + S_B + S_{A\times B} + S_e \tag{16.11}$$

其中

$$S_T=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{r}(x_{ijk}-\bar{x})^2$$

$$S_A=mr\sum_{i=1}^{n}(\bar{x}_{i\cdot\cdot}-\bar{x})^2$$

$$S_B=nr\sum_{j=1}^{m}(\bar{x}_{\cdot j\cdot}-\bar{x})^2$$

$$S_{AB}=r\sum_{i=1}^{n}\sum_{j=1}^{m}(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x})^2$$

$$S_e=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{r}(x_{ijk}-\bar{x}_{ij\cdot})^2$$

$$\bar{x}_{ij\cdot}=\frac{1}{r}\sum_{k=1}^{r}x_{ijk}$$

$$\bar{x}_{i\cdot\cdot}=\frac{1}{m}\sum_{j=1}^{m}\bar{x}_{ij\cdot}=\frac{1}{rm}\sum_{j=1}^{m}\sum_{k=1}^{r}x_{ijk}$$

$$\bar{x}_{\cdot j\cdot}=\frac{1}{n}\sum_{i=1}^{n}\bar{x}_{ij\cdot}=\frac{1}{rn}\sum_{i=1}^{n}\sum_{k=1}^{r}x_{ijk}$$

$$\bar{x}_{\cdots}=\bar{x}=\frac{1}{mnr}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{r}x_{ijk}$$

与上一节中的分析相仿，可以证明在假设 H_0(16.10)成立的条件下，$S_T/\sigma^2 \sim \chi^2(mnr-1)$，$S_A$ 的秩为 $f_A=n-1$，S_B 的秩为 $f_B=m-1$，S_{AB}的秩为 $f_{AB}=nm(r-1)$，S_e 的秩为 $f_e=(n-1)(m-1)$.由于

$$f_T = f_A + f_B + f_{AB} + f_e$$

根据定理 16.2.2 可知：S_A，S_B，S_{AB}，S_e 相互独立，而且 S_A/σ^2，S_B/σ^2，S_{AB}/σ^2，S_e/σ^2 分别服从自由度为 f_A，f_B，f_{AB}，f_e 的 χ^2 分布.因此

$$F_A=\frac{S_A/f_A}{S_e/f_e} \sim F(f_A,f_e) \tag{16.12}$$

$$F_B=\frac{S_B/f_B}{S_e/f_e} \sim F(f_B,f_e) \tag{16.13}$$

$$F_{AB}=\frac{S_{AB}/f_{AB}}{S_e/f_e} \sim F(f_{AB},f_e) \tag{16.14}$$

例 16.3.4 在某橡胶配方中，考虑 3 种不同的促进剂，4 种不同份量的氧化锌.同样的配方重复一次，测定 300%定强如 16.15 表所示.现在问：氧化锌，促进剂以及它们的交互作用对定强有无显著的影响.

表 16.15

氧化锌 B / 促进剂 A	B_1	B_2	B_3	B_4
A_1	31，33	34，36	35，36	39，38
A_2	33，34	36，37	37，39	38，41
A_3	35，37	37，38	39，40	42，44

(1) 列计算表格，如 16.16 表所示.

为了计算方便在计算时将 16.15 表所有数据同时减去 37，仍以 x_{ijk}表示.

表16.16的最后三列是经计算得到的，$X_{ij}=\sum_{k=1}^{r}x_{ijk}$，在目前的情况下$r=2$，所以$X_{ij}$是前两列之和，例如$X_{11}=-6-4=-10$；$\sum_{k=1}^{r}x_{ijk}^2$是前列数据的平方和，例如$\sum_{k=1}^{2}x_{11k}^2=(-6)^2+(-4)^2=36+16=52$；最后一列是$X_{ij}$的平方.

表　16.16

A	B	x_{ij1}	x_{ij2}	X_{ij}	$\sum_{k=1}^{2}x_{ijk}^2$	X_{ij}^2
A_1	B_1	−6	−4	−10	52	100
	B_2	−3	−1	−4	10	16
	B_3	−2	−1	−3	5	9
	B_4	2	1	3	5	9
A_2	B_1	−4	−3	−7	25	49
	B_2	−1	0	−1	1	1
	B_3	0	2	2	4	4
	B_4	1	4	5	17	25
A_3	B_1	−2	0	−2	4	4
	B_2	0	1	1	1	1
	B_3	2	3	5	13	25
	B_4	5	7	12	74	144
$\sum$				1	211	387

(2) 按下述公式进行计算：

n：因子A的水平数，

m：因子B的水平数，

r：试验的重复次数.

$$
\begin{cases}
P = \dfrac{1}{nmr}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}X_{ij}\right)^2 \\
Q = \dfrac{1}{mr}\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{m}X_{ij}\right)^2 \\
R = \dfrac{1}{nr}\sum\limits_{j=1}^{m}\left(\sum\limits_{i=1}^{n}X_{ij}\right)^2 \\
T = \dfrac{1}{r}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}X_{ij}^2 \\
W = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}\sum\limits_{k=1}^{r}x_{ijk}^2 \\
S_{总} = W - P, \qquad f_{总} = nmr - 1 \\
S_A = Q - P, \qquad f_A = n - 1 \\
S_B = R - P, \qquad f_B = m - 1 \\
S_{AB} = T - Q - R + P, \qquad f_{AB} = (n-1)(m-1) \\
S_e = W - T, \qquad f_e = nm(r-1) \\
F_A = \dfrac{S_A/f_A}{S_e/f_e} \\
F_B = \dfrac{S_B/f_B}{S_e/f_e} \\
F_{AB} = \dfrac{S_{AB}/f_{AB}}{S_e/f_e}
\end{cases}
$$

可得

$$P = \frac{1}{nmr}\cdot 1^2 = \frac{1}{3\times 4\times 2}1^2 \doteq 0$$

$$W = 211$$

$$T = \frac{1}{r}\times 387 = \frac{1}{2}\times 387 = 193.5$$

将 X_{ij}列成表 16.17.

表 16.17

	B_1	B_2	B_3	B_4	$\sum$	$(\sum)^2$
A_1	−10	−4	−3	3	−14	196
A_2	−7	−1	2	5	−1	1
A_3	−2	1	5	12	16	256
$\sum$	−19	−4	4	20		453
$(\sum)^2$	361	16	16	400	793	

由 16.17 表右下角的两个数得

$$Q = \frac{1}{mr} \cdot 453 = \frac{1}{4 \times 2} \cdot 453 = 56.6$$

$$R = \frac{1}{nr} \cdot 793 = \frac{1}{3 \times 2} \cdot 793 = 132.2$$

于是

$$S_T = W - P = 211$$

$$S_A = Q - P = 56.6$$

$$S_B = R - P = 132.2$$

$$S_{AB} = T - Q - R + P = 4.7$$

$$S_e = W - T = 17.5$$

$$F_A = \frac{56.6/2}{17.5/12} = \frac{28.3}{1.46} = 19.4$$

$$F_B = \frac{132.2/3}{17.5/12} = \frac{44.1}{1.46} = 30.2$$

$$F_{AB} = \frac{4.7/6}{17.5/12} = \frac{0.8}{1.46} = 0.55$$

上述计算结果可以列成方差分析表如表 16.18 所示.

表 16.18 有重复双因子方差分析表

方差来源	平方和	自由度	均方	F	显著性
A	$S_A=56.6$	$n-1=2$	28.3	19.4	* *
B	$S_B=132.2$	$m-1=3$	44.1	30.2	* *
$A\times B$	$S_{AB}=4.7$	$(n-1)(m-1)=6$	0.8	0.55	
误差	$S_e=17.5$	$nm(r-1)=12$	1.46		
总和	$S_T=211.0$	$nmr-1=23$			

(3) 统计检验

根据显著性水平 $\alpha(\alpha=0.01)$及相应的自由度查得 $F_{0.01}(2,12)=6.9$，$F_{0.01}(3,12)=6.0$，$F_{0.05}(6,12)=3.0$. 由于

$$F_A = 19.4 > 6.9$$

$$F_B = 30.2 > 6.0$$

$$F_{AB} = 0.55 < 3.0$$

可以认为促进剂 A 与氧化锌总量 B 都是重要因素；但是二者的交互作用很小，作用不显著.

16.4 均值的多重比较

在前几节，我们利用 F 检验判定因子的显著性. 当因子的水平数≥3 时，如果 F 检验显著，并不一定能说明所有水平之间的差异也显著，而只能说明至少存在着两个水平有显著差异. 在实际问题中，有时希望弄清楚哪些水平之间有显著差异，哪些水平之间不显著. 例如在例 16.2.4 中，5 种染整工艺对缩水率有显著影响，从平均缩水率看，以工艺 A_3 最低，但工艺 A_2 在生产上比较麻烦，劳动量比较大，自然想换成另一种与工艺 A_2 缩水率差异不显著的

工艺，后者在生产上却易于实施. 判定诸水平之间差异是否显著的方法叫**均值的多重比较**(multiple comparasions).

问题的一般提法是，已给 k 个平均值，按大小对它排序：$\bar{x}_{(1)} \geqslant \bar{x}_{(2)} \geqslant \cdots \geqslant \bar{x}_{(k)}$. 在做均值的成对多重比较时，如果均值的差 $\bar{x}_{(i)} - \bar{x}_{(j)}$ 大于某个临界值 D_a，那么将以显著性水平 α 断定水平 A_i 与 A_j 之间有显著差异. 选择 D_a 的方法有下述几种.

方法 16.4.1 多重比较方法

对于相等的，或不等的样本群以及任意的均值对，有

$$D_l = \sqrt{\bar{S}_e\left(\frac{1}{n_i} + \frac{1}{n_j}\right)(k-1)F_a(k-1, f_e)} \qquad (16.15)$$

其中

$\bar{S}_e = \dfrac{S_e}{f_e}$是方差分析中误差均值，

n_i, n_j 是被比较的两个水平的样本数.

f_e 是 S_e 的自由度，

k 是被比较的均值的个数，

α 是显著性水平，通常取 5%.

计算 $|\bar{x}_i - \bar{x}_j|$，若 $D_l > |\bar{x}_i - \bar{x}_j|$，则认为 $a_i \neq a_j$，否则认为 $a_i = a_j$.

方法 16.4.2 学生氏方法(或称 T 方法)

在样本群的大小相等时，有

$$D_1 = q(k, f_e)\sqrt{\frac{\bar{S}_e}{n}} \qquad (16.16)$$

当 $n_i \neq n_j$ 时，近似有公式

$$D'_1 = q(k, f_e)\sqrt{\bar{S}_e \times 0.5\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad (16.17)$$

其中 n 为每个水平重复试验的次数；k 为被比较的水平个数；

$\overline{S_e}$ 为方差分析中的均方误差；f_e 为 S_e 的自由度；$q(k,f_e)$是系数，可以从 q 表(附表 5)中查到.

例 16.4.3 设 $\alpha=0.05$，$\overline{x}_1$ 到 $\overline{x}_6$ 分别为：26.8，26.3，25.2，19.8，14.3，11.8. 又知 $n_i=8(i=1,2,\cdots,6)$，$\overline{S_e}=10.38$，$f_e=42$. 取 $\alpha=0.05$，由于 $k=6$，由 F 表查得

$$F_{0.05}(6-1,42)=2.44.$$

由此得

$$D_1=\sqrt{10.38\left(\frac{1}{8}+\frac{1}{8}\right)(6-1)2.44}=5.63.$$

以及由 q 表查得

$$q_{f_e}^{k}(\alpha)=q_{42}^{6}(0.05)=4.22.$$

由此

$$D_{3_{(6,0.05)}}=4.22\times\sqrt{\frac{10.38}{8}}=4.81.$$

计算均值差

$|\overline{x}_1-\overline{x}_2|=0.5$， $|\overline{x}_1-\overline{x}_3|=1.6$

$|\overline{x}_1-\overline{x}_4|=7.0$， $|\overline{x}_1-\overline{x}_5|=12.5$

$|\overline{x}_1-\overline{x}_6|=15.0$， $|\overline{x}_2-\overline{x}_3|=1.1$

$|\overline{x}_2-\overline{x}_4|=6.5$， $|\overline{x}_2-\overline{x}_5|=12.0$

$|\overline{x}_2-\overline{x}_6|=14.5$， $|\overline{x}_3-\overline{x}_4|=5.4$

$|\overline{x}_3-\overline{x}_5|=10.9$， $|\overline{x}_3-\overline{x}_6|=13.4$

$|\overline{x}_4-\overline{x}_5|=5.5$， $|\overline{x}_4-\overline{x}_6|=8.0$

$|\overline{x}_5-\overline{x}_6|=2.5$.

根据 D_1：

例如：$|\overline{x}_1-\overline{x}_2|=0.5<5.63$，可以接受 $a_1=a_2$，即认为水平 1

与水平 2 无显著差异. $|\bar{x}_1-\bar{x}_5|=12.5>5.63$,可以拒绝 $a_1=a_5$,即可认为水平 1 与水平 5 之间有显著差异.

根据 D_1:

例如:$|\bar{x}_1-\bar{x}_2|=0.5<4.81$,可以接受 $a_1=a_2$,可认为水平 1 与水平 2 无显著差异. $|\bar{x}_2-\bar{x}_5|=12.0>4.81$,可以拒绝 $a_2=a_5$,认为水平 2 与水平 5 之间有显著差异.

17 正交设计

17.1 正交表及其用法

正交设计(orthogonal design)是一种安排和分析试验的方法,它有下面两个主要特点:

(1) 任意一对因子(也称因素)的任一水平组合必在试验中出现,且出现次数相同.

(2) 总试验次数比用全面试验(所有因子的任一水平组合都进行搭配)要少许多次.例如对 7 个 2 水平因子,进行全面试验要做 $2^7=128$ 次.而用正交表安排试验只需做 8 次.

用正交表合理地安排试验,可以做到省时、省力、省钱,又能得到基本令人满意的检验效果.因此这种方法在改进产品质量、研究采用新工艺、试制新产品,了解设备工艺性能以及改进技术管理等众多方面都已获得应用.

17.1.1 正交表

一般形式的正交表(orthogonal table)可记为 $L_n(m_1^{k_1}\times m_2^{k_2}\times\cdots\times m_p^{k_p})$,它表示要做 n 次试验,可容纳水平数为 m_1 的 k_1 个因子,水平数为 m_2 的 k_2 个因子,…,以及水平数为 m_p 的 k_p 个因子.

例 17.1.1 正交表 $L_9(3^4)$如表 17.1.

常见的正交表中,如

2 水平的有 $L_4(2^3)$,$L_8(2^7)$,$L_{12}(2^{11})$,$L_{16}(2^{15})$,等;

3 水平的有 $L_9(3^4)$,$L_{27}(3^{13})$,等;

4 水平的有 $L_{16}(4^5)$,等;

5 水平的有 $L_{25}(5^6)$，等.

表 17.1

试验号 \ 列号	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

正交表有下面两个特点：

(1) 表中每列中不同数字出现的次数是相等的. 如 $L_9(3^4)$中，每列中 1,2,3 数字各出现 3 次.

(2) 在任意两列中，将同一行的两个数字看成有序数对时，每种数对出现的次数是相等的，如 $L_9(3^4)$表中，每两列中每行的有序数对有(1,1)，(1,2)，(1,3)，(2,1)，(2,2)，(2,3)，(3,1)，(3,2)，(3,3)等 9 个. 它们各出现一次.

由于正交表有这两条性质，所以用它来安排试验时，各因子的各种水平搭配是均衡的.

17.1.2 正交表的使用方法

例 17.1.2 某炼铁厂为了提高铁水温度，需要通过试验选择最好的生产方案，经初步分析主要有 3 个因子影响铁水温度. 它们

是焦比、风压和底焦高度. 每个因子考虑 3 个水平. 问这 3 个因子、3 个水平如何安排才能获得最高的铁水温度.

表 17.2

水平 \ 因子	焦比(A)	风压(B) 133Pa	底焦高度(C) m
1 水平	1 : 16	22610	1.2
2 水平	1 : 18	30590	1.5
3 水平	1 : 14	26600	1.3

在此问题中我们关心的是铁水温度，称铁水温度为试验指标. 由于只有一个指标故称为单指标问题，如果做全面试验需做 3^3 次试验，用正交表 $L_9(3^4)$安排试验，只需做 9 次试验. 利用 $L_9(3^4)$表，将 A,B,C 这 3 个因子安排在表的前 3 列上，按照表 17.2 所确定的水平及 $L_9(3^4)$中的试验号进行试验(例如将因子 A 的第 1 个水平 1 : 16 与因子 B 的第 1 个水平 22610Pa 及因子 C 的第 1 个水平 1.2m 高的底焦搭配做一次试验称为标号为 1 的试验)，得结果如表 17.3 所示.

正交试验结果的分析

每列中 $K_1,K_2,K3$ 的数字是这样得到的:将表中第 1 列中数字 1 所对应的指标中的数字加在一起得到第 1 列中 K_1 的数字，例如将表中第 1 列中 1 所对应的指标数字 15,45,35(表中最右一列)相加得到 95，写在第 1 列下 K_1 的那一行中，再将表中第 2 列中 1 所对应的指标数字 15,40,40 相加，得到 95，写在第 2 列下 K_1 的那一行中，最后再将表中第 3 列中 1 所对应的指标数字 15,30,40 相加，得到 85，写在第 3 列下 K_1 的那一行中，…，表中 K_3 那一行的三个数字，也是分别由第 1 列中数字 3(表示第 3 个水平)，第 2 列中的数字 3，以及第 3 列中的数字 3 所对应的指标数字相加后

记录在 K_3 的那一行,它们分别是 140,155,120.

表 17.3

列号 / 试验号	1 A	2 B	3 C	指标 铁水温度(℃)	铁水温度 减去 1350
1	1	1	1	1365	15
2	1	2	2	1395	45
3	1	3	3	1385	35
4	2	1	2	1390	40
5	2	2	3	1395	45
6	2	3	1	1380	30
7	3	1	3	1390	40
8	3	2	1	1390	40
9	3	3	2	1410	60
K_1	95	95	85		
K_2	115	130	145		
K_3	140	155	120		
k_1	31.7	31.7	28.3		
k_2	38.3	43.3	48.3		
k_3	46.7	41.7	40.0		
极差	15.0	11.6	20.0		
优方案	A_3	B_2	C_2		

表中的 k_1,k_1,k_3 行的数字是这样得到的:例如第 1 列中 k_1 行的数字31.7 是由第 1 列 K_1 行的数字 95 除以 3 得来,因为第 1 列中1 出现 3 次,故而将其总和除以第 1 列中 1 出现的次数 3.

由于每列中 1,2,3 出现的次数都是 3. 因此 $k_1=K_1/3$, $k_2=K_2/3$, $k_3=K_3/3$.

表中极差那一行的数字是由每列中 k_1,k_2,k_3 中的最大值减去最小值得到. 例如 $15.0=46.7-31.7$.

极差最大者,表示该因素的各个水平的变化对指标影响大,因而该因子为主要因子,从这个试验看出 C 因子(底焦高度)的极差最大,而且 $k_2(=48.3)$ 在 k_1,k_2,k_3 中为最大,故而对指标好的水平选为 C_2, A 因子(焦比)极差次之,而以 $k_3(=46.7)$ 最大,故而选第 3 水平 A_3(焦比 1∶14).

B 因子(风压)极差最小,其中又以 $k_2(=43.3)$ 最大,选 B_2 (30590 Pa). 最后得

优方案: $A_3C_2B_2$. 如果不放心,可再做一次试验 $A_3C_2B_2$,观察其结果是否是最佳方案.

方法 17.1.3 用正交表安排试验的步骤

(1) 明确试验目的,确定要考核的试验指标.

(2) 确定要考察的因子和因子的水平.

(3) 选用合适的正交表,能安排下因子和水平即可.

(4) 根据试验号的安排进行试验并纪录试验指标的具体数据.

(5) 计算 $K_1,K_2,K_3,\cdots$(K_i 为对应于第 i 个水平所对应的指标总和).

(6) 计算 $k_1,k_2,k_3,\cdots$,

$k_i=K_i/$该列中第 i 个水平出现的次数.

(7) 计算每列中 $k_1,k_2,\cdots$的极差.

(8) 极差最大的列所对应的因子是最主要因子. 当要求指标愈高愈好时可选该列中 $k_1,k_2,\cdots$,中最大者所对应的水平为优水平. 当要求指标愈低愈好时,可选该列中 $k_1,k_2,\cdots$中最小者所对应之水平为优水平.

每列都进行分析.从而得到最优的试验安排.

17.2 多指标的分析方法

17.1的问题中试验指标只有一个，考察起来较方便，但在实际问题中需要考虑的指标往往不止一个.下面介绍两种解决多指标的试验方法.

17.2.1 综合平衡法

例 17.2.1 对某产品的原料进行配方试验，要检验3项指标：抗压强度、落下强度（指产品从高处落下时出现破裂现象的最小高度）和裂纹度.前两个指标愈大愈好，第3个指标越小越好.根据以往经验，配方中有三个重要因子：水份、粒度和碱度.它们各考虑3个水平，如表17.4所示.

表 17.4

因子 \ 水平	水份(%) A	粒度(%) B	碱度
1	8	4	1.1
2	9	6	1.3
3	7	8	1.5

选用 $L_9(3^4)$表，将因子 A，B，C 放在前3列中.

对3个指标分别按方法17.1.3中所述进行计算，结果列在表17.5的下端.

从结果看出，某个水平组合对某个指标优良而对另一个指标就不一定优良，故而需要做综合分析.

表 17.5

试验号＼列号		1 A	2 B	3 C	抗压强度 (kg/个)	落下强度 (0.5m/次)	裂纹度
1		1	1	1	11.5	1.1	3
2		1	2	2	4.5	3.6	4
3		1	3	3	11.0	4.6	4
4		2	1	2	7.0	1.1	3
5		2	2	3	8.0	1.6	2
6		2	3	1	18.5	15.1	0
7		3	1	3	9.0	1.1	3
8		3	2	1	8.0	4.6	2
9		3	3	2	13.4	20.2	1
抗压强度	K_1	27.0	27.5	38.0			
	K_2	33.5	20.5	24.9			
	K_3	30.4	42.9	28.0			
	k_1	9.0	9.2	12.7			
	k_2	11.2	6.8	8.3			
	k_3	10.1	14.3	9.3			
	极差	2.2	7.5	4.4			
	优方案	A_2	B_3	C_1			

续表

试验号＼列号		1 A	2 B	3 C	各指标试验结果 抗压强度(kg/个)	落下强度(0.5m/次)	裂纹度
落下强度	K_1	9.3	3.3	20.8			
	K_2	17.8	9.8	24.9			
	K_3	25.9	39.9	7.3			
	k_1	3.1	1.1	6.9			
	k_2	5.9	3.3	8.3			
	k_3	8.6	13.3	2.4			
	极差	5.5	12.2	5.9			
	优方案	A_3	B_3	C_2			
裂纹度	K_1	11	9	5			
	K_2	5	8	8			
	K_3	6	5	9			
	k_1	3.7	3.0	1.7			
	k_2	1.7	2.7	2.7			
	k_3	2.0	1.7	3.0			
	极差	2.0	1.3	1.3			
	优方案	A_2	B_3	C_1			

将 3 个因子对 3 个指标的影响各画出一个图.

从表 17.5 看出欲使抗压强度高,优方案为 $A_2B_3C_1$,欲使落下

强度高，优方案为$A_3B_3C_2$，欲使裂纹度低，优方案为$A_2B_3C_1$. 通过对图 17.1～图 17.3 进行综合分析，可以看出对于因子 A 选 A_2 可

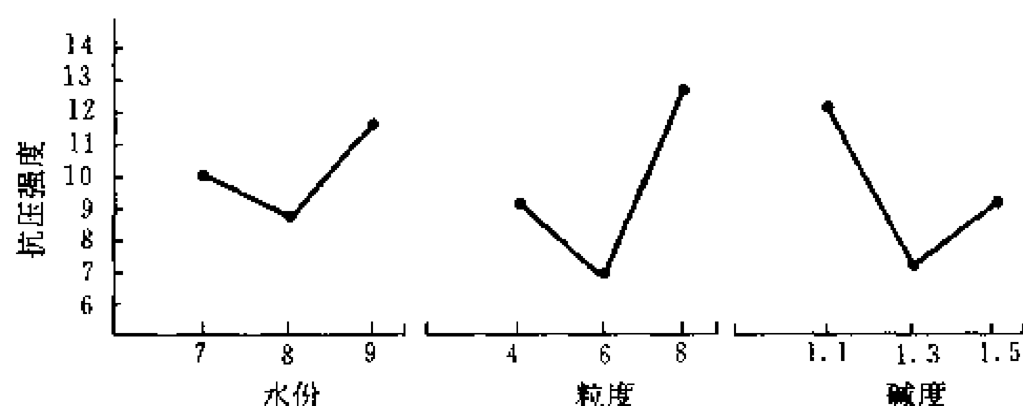


图 17.1

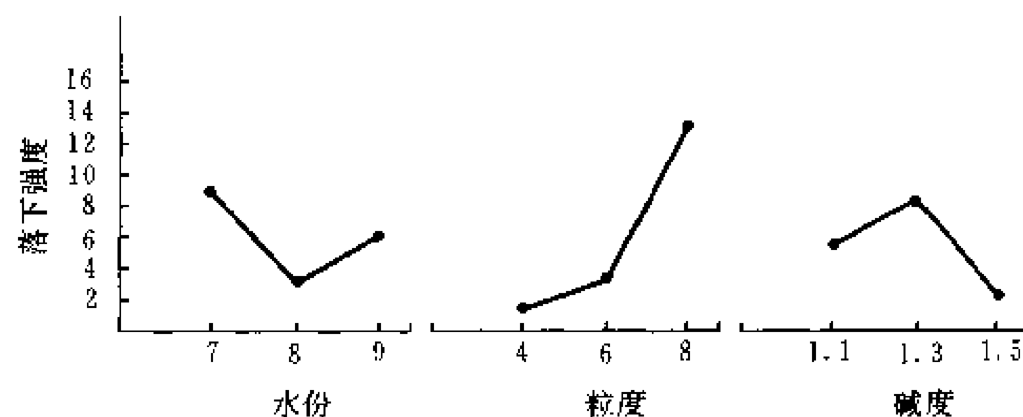


图 17.2

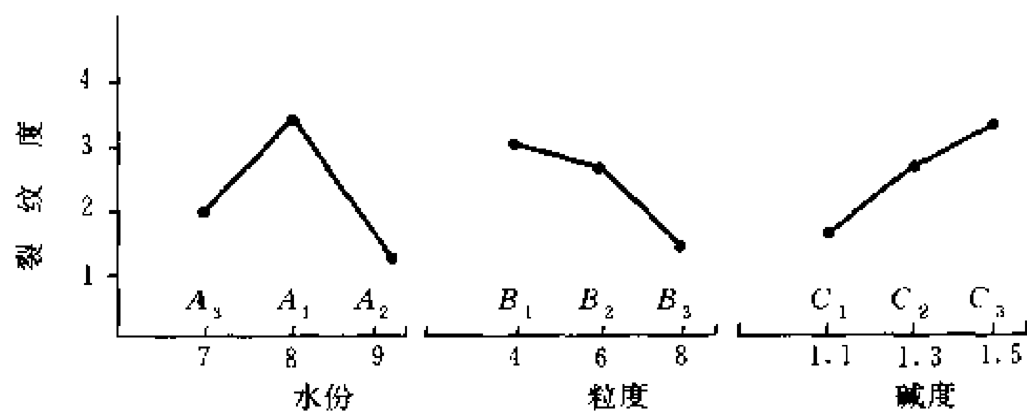


图 17.3

使抗压强度高且裂纹度低,选 B_3 可使 3 个指标达到好的要求,选 C_1 也可使抗压强度高且裂纹度低.因此综合分析的结果说明优方案应为 $A_2B_3C_1$.

17.2.2 综合评分法

例 17.2.2 某厂生产一种化工产品,需要检验两个指标:核酸纯度与回收率.这两个指标都是越大越好.考虑 4 个因子,各 3 个水平如表 17.6.

表 17.6

因子 / 水平	时间(h) A	加料中核酸含量 B	PH 值 C	加水量 D
1	25	7.5	5.0	1∶6
2	5	9.0	6.0	1∶4
3	1	6.0	9.0	1∶2

选正交表 $L_9(3^4)$ 如表 17.7.

综合评分法就是根据各个指标的重要性的不同,给每一个试验评出一个分数.如在 17.2.2 例中,根据实际情况纯度的重要性比回收率要重要得多.可认为纯度的重要性是回收率的 4 倍,即若将回收率的重要性看成 1,纯度就是 4,这个 1 和 4 分别叫做两个指标的权.按这个权给每个试验打总分:

$$总分=4\times纯度+1\times回收率.$$

根据这个公式,算出每个试验的总分.列在表 17.7 的最右端.然后对这个综合评分做分析.根据极差的大小以及总分越高越好的要求,得优方案是 $A_1B_3C_2D_1$.

表 17.7

试验号＼列号	1 A	2 B	3 C	4 D	各指标试验结果 纯度	回收率	综合评分
1	1	1	1	1	17.5	30.0	100.0
2	1	2	2	2	12.0	41.2	89.2
3	1	3	3	3	6.0	60.0	84.0
4	2	1	2	3	8.0	24.2	56.2
5	2	2	3	1	4.5	51.0	69.0
6	2	3	1	2	4.0	58.4	74.4
7	3	1	3	2	8.5	31.0	65.0
8	3	2	1	3	7.0	20.5	48.5
9	3	3	2	1	4.5	73.5	91.5
K_1	273.2	221.2	222.9	260.5			
K_2	196.6	206.7	236.9	228.6			
K_3	205.0	249.9	218.9	188.7			
k_1	91.1	73.7	74.3	86.8			
k_2	65.5	68.9	79.0	76.2			
k_3	68.3	83.9	72.7	62.9			
极差	25.6	14.4	6.3	23.9			
优方案	A_1	B_3	C_2	D_1			

17.3 混合水平的正交设计

在实际问题中各因子的水平数不一定相等，解决这样的问题有两种方法：

(1) 直接用混合水平的正交表.

(2) 拟水平法,即把水平不同的问题转化成水平相同的问题来处理.

17.3.1 混合水平的正交表及其用法

混合水平的正交表有许多种,其中如 $L_8(4^1\times 2^4)$就是一种,它能解决一个因子有 4 个水平,及 4 个因子各有 2 个水平的试验.

例 17.3.1 某农科站进行品种试验,共考虑 4 个因子,一个因子有 4 个水平,另外 3 个因子各有 2 个水平. 试验指标是产量,越高越好,见表 17.8.

表 17.8

因子 水平	品种 A	氮肥量 B	氮、磷、钾肥比例 C	规格 D
1	甲	25	3∶3∶1	6×6
2	乙	30	2∶1∶2	7×7
3	丙			
4	丁			

选择正交表 $L_8(4^1\times 2^4)$,将 A,B,C,D 因子放在前 4 列如表 17.9.

17.3.2 拟水平法

例 17.3.2 做某试验,试验指标只有一个,它的数值越小越好. 在 4 个因子中 A,B,D 是 3 水平的,C 因子是 2 水平的(表 17.10).

如果 C 因子也是 3 水平的,那么可选用 $L_9(3^4)$来安排试验. 但现在 C 是 2 水平的,为了能使用 $L_9(3^4)$表,我们将因子 C 虚拟

一个水平成为 3 个水平. 使用 $L_9(3^4)$表(表 17.11).

表 17.9

因子 / 试验号	A	B	C	D	试验指标(产量)	产量减去 200
1	1	1	1	1	195	-5
2	1	2	2	2	205	5
3	2	1	1	2	220	20
4	2	2	2	1	225	25
5	3	1	2	1	210	10
6	3	2	1	2	215	15
7	4	1	2	2	185	−15
8	4	2	1	1	190	−10
K_1	0	10	20	20		
K_2	45	35	35	25		
K_3	25					
K_4	−25					
k_1	0	2.5	5.0	5.0		
k_2	22.5	8.8	6.3	6.3		
k_3	12.5					
k_4	−12.5					
极差	35.0	6.3	1.3	1.3		
优方案	A_2	B_2	C_2	D_2		

表 17.10

水平 \ 因子	A	B	C	D
1	350	15	60	65
2	250	5	80	75
3	300	10	[80]	85

虚线框内就是虚拟的水平.

表 17.11

试验号 \ 因子	A	B	C		D	试验指标
1	1	1	1	1	1	45
2	1	2	2	2	2	36
3	1	3	3	2	3	12
4	2	1	2	2	3	15
5	2	2	3	2	1	40
6	2	3	1	1	2	15
7	3	1	3	2	2	10
8	3	2	1	1	3	5
9	3	3	2	2	1	47
K_1	93	70		65	132	
K_2	70	81		160	61	
K_3	62	74			32	
k_1	31.0	23.3		21.7	44.0	
k_2	23.3	27.0		26.7	20.3	
k_3	20.7	24.7			10.7	

续表

试验号 \ 因子	A	B	C	D	试验指标
$k_2=\frac{K_2}{6}$					
极差	10.3	3.7	5.0	33.3	
优方案	A_3	B_1	C_1	D_3	

由于用正交表做试验时因子 C 的第 3 水平就是第 2 水平，因此这一列的真正水平写在虚线框内．对于因子 C 的 K 值，也只出现 $K_1,K_2,k_1=K_1/3,k_2=K_2/6$，而对于因子 A,B,D 列的 k_1,k_2,k_3，都是 $k_i=K_i/3,i=1,2,3$，优方案是 $A_3B_1C_1D_3$．

17.4 有交互作用的正交设计

例 17.4.1 有 4 块试验田，土质情况基本一样，种植同样的作物，现将氮肥、磷肥采用不同方式施在四块田里，收获后算出平均亩产．

表 17.12 单位：kg

氮肥 \ 磷肥	$P_1=0$	$P_2=4$
$N_1=0$	400	450
$N_2=6$	430	560

从表 17.12 看出，不加化肥，平均亩产 400(kg)．只加磷肥(4kg)，平均收获 450(kg)，增产 50kg．只施加氮肥(6kg)收获 430(kg)，增产 30kg．而氮磷肥都施增产 160kg，因此氮、磷肥的交互作用显然可用下面公式：

交互作用＝氮、磷肥的总效果－（只加氮肥的效果＋只加磷肥的效果）.

在多因子试验中，如果不能确认交互作用影响很小，就应该通过试验分析交互作用.

17.4.1 交互作用表

它是用来安排交互作用试验的，我们把交互作用看成一个因子，让它也占表中的一列，叫交互作用列. 交互作用的安排可以查交互作用表.

表 17.13

列号（ ）＼列号	1	2	3	4	5	6	7
	(1)	3	2	5	4	7	6
		(2)	1	6	7	4	5
			(3)	7	6	5	4
				(4)	1	2	3
					(5)	3	2
						(6)	1
							(7)

从表 17.13 可以查出正交表 $L_8(2^7)$中的任何两列的交互作用列，查法如下：第 1 个列号是带（ ）的，从左往右看第 2 个列号是不带（ ）的，从上往下看. 表中交叉处的数字就是两列的交互作用列的列号. 例如用 $L_8(2^7)$安排试验时，如果因子 A 放在第 1 列，因子 B 放在第 2 列，则从带括号的(1) 往右看，从不带括号的 2 往下看，交叉处的数字是 3，这就是 A 因子与 B 因子的交互作用列. 从表 17.13 还可看出，第 1 列和第 3 列的交互作用列号是第 2 列，第 4 列与第 6 列的交互作用列也是第 2 列. 这说明不同列的交互作用列可能在同一列.

17.4.2 正交表的列的自由度

(1) 正交表每列的自由度 $f_{列}$ 等于各列的水平数减 1. 由于因子与列是等同的，因而每个因子的自由度等于该因子的水平数减 1.

(2) 两因子交互作用的自由度等于两因子的自由度乘积.

$$f_{A\times B}=f_A\times f_B$$

一般原则：两个 n 水平的因子，由于每个因子的自由度为$(n-1)$，所以这两个因子的交互作用的自由度为$(n-1)\times(n-1)$，此时交互作用应占$(n-1)\times(n-1)$列.

17.4.3 水平数相同的有交互作用的正交设计

例 17.4.2 某产品的产量取决于 3 个因子 A,B,C，每个因子都有两个水平(表 17.14). 试验指标为产量，越高越好.

表 17.14

水平 \ 因子	A	B	C
1	60	1.2	20%
2	80	1.5	30%

这是 3 因子，2 水平的问题，要考虑交互作用，利用表 17.13，3 个因子 A,B,C 要占 3 列，它们之间的交互作用 $A\times B,A\times C,B\times C$ 又占 3 列，共占 6 列，可用正交表 $L_8(2^7)$安排试验. 若将 A,B 分别放在第 1,2 列，从表 17.13 查出 $A\times B$ 应在第 3 列，因此 C 就不能放在第 3 列. 现将 C 放在第 4 列，由表 17.13 查出 $A\times C$ 应在第 5 列，$B\times C$ 应在第 6 列. 用直观分析法把交互作用当成新的因子看待. 当然在做试验时仍用 A,B,C 的在正交表中的水平搭配作试验，而在 $A\times B,A\times C,B\times C$ 等列中的各水平只是用于分析试

验结果的(表 17.15).

表　17.15

列号 试验号	1 A	2 B	3 $A\times B$	4 C	5 $A\times C$	6 $B\times C$	指标
1	1	1	1	1	1	1	65
2	1	1	1	2	2	2	73
3	1	2	2	1	1	2	72
4	1	2	2	2	2	1	75
5	2	1	2	1	2	1	70
6	2	1	2	2	1	2	74
7	2	2	1	1	2	2	60
8	2	2	1	2	1	1	71
K_1	285	282	269	627	282	281	
K_2	275	278	291	293	278	279	
k_1	142.5	141.0	134.5	133.5	141.0	140.5	
k_2	137.5	139.0	145.5	146.5	139.0	139.5	
极差	5	2	11	13	2	1	
优方案	A_1	B_1	2 水平	C_2	1 水平	1 水平	
优方案	A_1	B_2		C_2			

分析结果:从极差看,影响最大的是 C 因子,以 2 水平最好.其次是 $A\times B$,也以 2 水平为好.再其次是因子 A,以 1 水平为好.影响较小的因子是 B 因子,其第 1 水平与第 2 水平差不多.因此,C 取第 2 水平,$A\times B$ 取第 2 水平,A 取第 1 水平,从表中看,此时 B 因子取第 2 水平的试验,也即第 4 号试验是个较优的方案.从 8 个试验中看,这个方案确是 8 个方案中最好的一个.

17.5 正交设计的方差分析

用正交表安排 m 个因子的试验，试验总次数为 n，试验结果(指标)为 $x_1,x_2,\cdots,x_n$. 假定每个因子为 n_a 个水平，每个水平在 a 次试验中出现，则 $n=an_a$.

17.5.1 离差平方和

公式 17.5.1 总离差平方和 S_T

$$S_T = W - P$$

其中

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$S_T = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2.$$

$$W = \sum_{i=1}^{n} x_i^2,$$

$$P = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2.$$

公式 17.5.2 各因子离差平方和

设 A 因子安排在第 l 列，用 x_{ij}表示因子 A 第 i 个水平的第 j 个试验的结果. $i=1,2,\cdots,n_a$; $j=1,2,\cdots,a$

则有

$$\sum_{i=1}^{n_a}\sum_{j=1}^{a} x_{ij} = \sum_{k=1}^{n} x_k,$$

$$S_A = Q_A - P,$$

其中

$$S_A = \frac{1}{a}\sum_{i=1}^{n_a}\left(\sum_{j=1}^{a} x_{ij}\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n_a}\sum_{j=1}^{a} x_{ij}\right)^2$$

$$= \frac{1}{a}\sum_{i=1}^{n_a} k_i^2 - \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right)^2.$$

$$Q_A = \frac{1}{a}\sum_{i=1}^{n_a} k_i^2,$$

$$k_i = \sum_{j=1}^{a} x_{ij}.$$

公式 17.5.3 试验误差的离差平方和 S_E

$$S_T = S_{因+交} + S_E$$

$$S_E = S_T - S_{因+交}$$

17.5.2 计算自由度

$f_{总}$＝试验总次数－1＝$n-1$.

各因子自由度：

$f_{因}$＝因子水平数－1＝n_a-1.

（如两因子交互作用自由度：$f_{A\times B} = f_A f_B$）

公式 17.5.4 试验误差自由度 f_E

$$f_E = f_{总} - f_{因+交}.$$

17.5.3 计算平均离差平方和

公式 17.5.5

$$因子平均离差平方和 = \frac{因子的离差平方和}{因子的自由度} = \frac{S_{因}}{f_{因}}.$$

公式 17.5.6

$$试验误差的平均离差平方和 = \frac{试验误差的离差平方和}{试验误差自由度} = \frac{S_E}{f_E}.$$

17.5.4 *F* 值的计算

公式 17.5.7

$$F_{因}=\frac{\text{因子的平均离差平方和}}{\text{试验误差的平均离差平方和}}\sim F(f_{因},f_E).$$

如果 $F_{因}>F_{\alpha}(f_{因},f_E)$，则认为该因子的影响显著. 如果 $F_{因}<F_{\alpha}(f_{因},f_E)$，则认为该因子的影响不显著.

例 17.5.8 为提高某产品的质量，需要考虑 3 个因素：反应温度，反应压力和溶液浓度，每个因素都取 3 个水平(表 17.16).

表 17.16

因素 / 水平	温度(℃) A	压力 B	浓度 C
1	60	2.0	0.5
2	65	2.5	1.0
3	70	3.0	2.0

考虑因素之间所有一级交互作用，要找出较好的工艺条件.

这是 3 因素 3 水平的试验，每两个因素的交互作用要占两列. 3 个因素的所有交互作用共占 6 列，加上因素 A,B,C，所以在正交表中共占 9 列. 应选正交表 $L_{27}(3^{13})$，并查该正交表的交互作用表，其表头设计如下：

表 17.17 $L_{27}(3^{13})$有交互作用的表头设计

列号 / 因素	1	2	3	4	5	6	7	8	9	10	11	12	13
	A	B	$(A\times B)_1$	$(A\times B)_2$	C	$(A\times C)_1$	$(A\times C)_2$	$(B\times C)_1$			$(B\times C)_2$		

现将正交表 $L_{27}(3^{13})$ 去掉 9，10，12，13 列，连同试验结果、分析计算过程列在表 17.18 中，此例中每个因子有 3 个水平，因此 $n_a=3$，每个水平下都做 9 次试验，因此 $a=9$. 总共做试验次数为 $n=an_a=9\times3=27$ 次.

表 17.18

试验号＼因子	A	B	$(A\times B)_1$	$(A\times B)_2$	C	$(A\times C)_1$	$(A\times C)_2$	$(B\times C)_1$	$(B\times C)_2$	指标 产量
1	1	1	1	1	1	1	1	1	1	1.30
2	1	1	1	1	2	2	2	2	2	4.63
3	1	1	1	1	3	3	3	3	3	7.23
4	1	2	2	2	1	1	1	2	3	0.50
5	1	2	2	2	2	2	2	3	1	3.67
6	1	2	2	2	3	3	3	1	2	6.23
7	1	3	3	3	1	1	1	3	2	1.37
8	1	3	3	3	2	2	2	1	3	4.73
9	1	3	3	3	3	3	3	2	1	7.07
10	2	1	2	3	1	2	3	1	1	0.47
11	2	1	2	3	2	3	1	2	2	3.47
12	2	1	2	3	3	1	2	3	3	6.13
13	2	2	3	1	1	2	3	2	3	0.33
14	2	2	3	1	2	3	1	3	1	3.40
15	2	3	3	1	3	1	2	1	2	5.80
16	2	3	1	2	1	2	3	3	2	0.63
17	2	3	1	2	2	3	1	1	3	3.97
18	2	3	1	2	3	1	2	2	1	6.50
19	3	1	3	2	1	3	2	1	1	0.03
20	3	1	3	2	2	1	3	2	2	3.40
21	3	1	3	2	3	2	1	3	3	6.80
22	3	2	1	3	1	3	2	2	3	0.57
23	3	2	1	3	2	1	3	3	1	3.97

续表

因子 / 试验号	A	B	$(A\times B)_1$	$(A\times B)_2$	C	$(A\times C)_1$	$(A\times C)_2$	$(B\times C)_1$	$(B\times C)_2$	指标 产量
24	3	2	1	3	3	2	1	1	2	6.83
25	3	3	2	1	1	3	2	3	2	1.07
26	3	3	2	1	2	1	3	1	3	3.97
27	3	3	2	1	3	2	1	2	1	6.57
K_1	36.73	33.46	35.63	34.30	6.27	32.94	34.21	33.33	32.98	100.64
K_2	30.70	31.30	32.08	31.73	35.21	34.66	33.13	33.04	33.43	
K_3	33.21	35.88	32.93	34.61	59.16	33.04	33.30	34.27	34.23	
极差	6.03	4.58	7.30	2.88	52.89	1.72	1.08	0.23	1.25	
优方案	A_1	B_3	$(A\times B)$的第1水平		C_3					

计算结果如下：

$$P=\frac{1}{27}(100.64)^2=375.13.$$

$$Q_A=\frac{1}{9}(36.73^2+30.70^2+33.21^2)=377.17.$$

$$Q_B=\frac{1}{9}(33.46^2+31.30^2+35.88^2)=376.29.$$

$$Q_C=\frac{1}{9}(6.27^2+35.21^2+59.16^2)=531.00.$$

$$Q_{(A\times B)_1}=\frac{1}{9}(35.63^2+32.08^2+32.93^2)=375.89.$$

$$Q_{(A\times B)_2}=\frac{1}{9}(34.30^2+31.73^2+34.61^2)=375.68.$$

$$Q_{(A\times C)_1}=\frac{1}{9}(32.94^2+34.66^2+33.04^2)=375.33.$$

$$Q_{(A\times C)_2}=\frac{1}{9}(34.21^2+33.13^2+33.30^2)=375.20.$$

$$Q_{(B\times C)_1}=\frac{1}{9}(33.33^2+33.04^2+34.27^2)=375.22.$$

$$Q_{(B\times C)_2}=\frac{1}{9}(32.98^2+33.43^2+34.23^2)=375.22.$$

再计算出：

$$S_A=Q_A-P=2.04.$$

$$S_B=Q_B-P=1.17.$$

$$S_C=Q_C-P=155.87.$$

$$S_{A\times B}=S_{(A\times B)_1}+S_{(A\times B)_2}=(Q_{(A\times B)_1}-P)+(Q_{(A\times B)_2}-P)$$
$$=Q_{(A\times B)_1}+Q_{(A\times B)_2}-2P=1.32.$$

$$S_{A\times C}=Q_{(A\times C)_1}+Q_{(A\times C)_2}-2P=0.28.$$

$$S_{B\times C}=Q_{(B\times C)_1}+Q_{(B\times C)_2}-2P=0.18.$$

总平方和 $W=\sum_{k=1}^{27}x_k^2=536.33.$

$$S_T=W-P=161.20.$$

$$S_E=S_T-S_{因\cdot交}$$
$$=S_T-(S_A+S_B+S_C+S_{A\times B}+S_{A\times C}+S_{B\times C})$$
$$=0.34.$$

计算自由度：

$$f_A=f_B=f_C=3-1=2.$$

$$f_{A\times B}=f_{A\times C}=f_{B\times C}=2\times 2=4.$$

$$f_{总}=n-1=27-1=26.$$

$$f_E=f_{总}-f_{因+交}=26-(2\times 3+4\times 3)=8.$$

计算平均离差平方和及 F 值并列表如下：

由于这个例题的试验指标是产量，当然越大越好．所以最优方案应取各因子中 K 的最大值所对应的水平．从表 17.19 中看出，因子 A 应取第 1 水平 A_1，因子 B 应取第 3 水平 B_3，因子 C 应取第 3 水平 C_3．

表 17.19

差异来源	离差平方和	自由度	平均离差平方和	F 值	临界值	显著性	优方案
A	2.04	2	1.02	20.40	$F_{0.01}(2,16)$	* *	A_1
B	1.17	2	0.58	11.60	$=6.23$	* *	B_3
C	155.87	2	77.93	1559.60	$F_{0.01}(4,16)$	* *	C_3
$A\times B$	1.32	4	0.33	6.60	$=4.77$		$A_1B_3C_3$
$A\times C$	0.28	4 }					
$B\times C$	0.18	4 } 16	0.05				
误差	0.34	8 }					
总和	161.20	26					

交互作用 $A\times B$ 也是显著的,但是由于 $A\times B$ 占两列,直观分析法有些困难,因此把 A 和 B 的各种组合试验结果对照起来分析如表 17.20 所示.

表 17.20

B \ A	1	2	3
1	13.16	10.07	10.23
2	10.40	9.53	11.37
3	13.17	11.10	11.61

从表 17.20 看出,当 A 取第 1 水平,B 取第 3 水平时试验结果为 13.17,是所有结果中最大值.这与单独考虑因子 A,B 时所得结果是一致的.所以优方案应为 $A_1B_3C_3$.

18 多元分析(Ⅰ)

多元分析(multivariate analysis)是一类处理多元数据统计方法的总称,是统计学中一个非常重要的,包含丰富理论成果和众多应用方法的分支.

在第11～13章中,着重讨论的是有关一元统计的概念和方法,其中待处理的变量只有一个.在第14章回归分析和第15章方差分析中,虽然涉及到多个变量,但总假定其中只有一个是随机变量,其余的变量都是可以精确测定的.但在实际问题中常要同时研究多个随机变量,这些随机变量间的相互关系和相互影响,这就要研究多元分析.

多元分析的内容比一元统计要丰富得多.一方面,某些一元统计的方法可以经过适当的推广应用于多元数据处理.例如多元回归分析与多元方差分析等.由于经过了推广,这些方法的内涵更丰富了.另一方面,多元数据本身引出了一些在一元数据处理方法中不曾提出的问题,例如研究多个变量之间的相互关系问题.于是,便产生了主成份分析、因子分析及典则相关分析等方法.

经典的多元分析理论是建立在数据服从正态分布的假定前提上的.这样的假定带来一系列理论分析上的好处.例如,不相关等价于独立性,条件期望具有线性形式及中心极限定理保证了在大样本的前提下正态性得以成立等.因此,在正态假定下建立起的多元统计方法具有理论上严谨及计算上以线性问题为主的优点.但是随着统计学的深入发展及实际问题提出不同的要求,一些新的多元统计方法产生了.这些方法通常不要求数据服从正态分布,并根据实际问题的需要给出不同的模型和灵活多变的计算方法,以

取得满意的实际效果. 但是,这些新方法的缺点是尚缺乏坚实的理论基础.

18.1 多元正态分布理论

18.1.1 多元正态分布

设 $\boldsymbol{x}$ 是 p 元随机变量,其均值向量为 $\boldsymbol{\mu}$,方差矩阵为 $\boldsymbol{\Sigma}$.

定义 18.1.1 若对任意 p 元常向量 $\boldsymbol{a}$,$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 服从一元正态分布,则称 $\boldsymbol{x}$ 服从 p 元正态分布,记作 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

注 1 在上述定义中包含了如下的假定:常数被看成服从退化的正态分布.

注 2 当方差矩阵 $\boldsymbol{\Sigma}$ 为半正定时,常称 $\boldsymbol{x}$ 服从退化的 p 元正态分布.

注 3 在不致引起混淆的场合,常将下标 p 省略,记作 $\boldsymbol{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

定理 18.1.2 设 $\boldsymbol{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$,则其特征函数可表为

$$\varphi(\boldsymbol{t}) = \exp\left\{i\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma},\boldsymbol{t}\right\},$$

其中 $\boldsymbol{t}^{\mathrm{T}}=(t_1,\cdots,t_p)$为任意 p 元实向量.

定理 18.1.3 设 $\boldsymbol{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$,$\boldsymbol{\Sigma}$ 正定,则 $\boldsymbol{x}$ 有密度函数

$$f(\boldsymbol{x}) = \left\{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\right\}^{-1}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

定理 18.1.4 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$,$\boldsymbol{A}$ 为一个 $q\times p$ 常数矩阵,$\boldsymbol{c}$ 为一个 q 元常向量,则

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{c} \sim N_q(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{c},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}),$$

推论 18.1.5 正态随机向量的任何子向量仍服从正态分布. 特别,设 $\boldsymbol{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$,则 $\boldsymbol{x}$ 的第 i 个分量 $x_i\sim N(\mu_i,\sigma_{ii})$,其中 μ_i 为 $\boldsymbol{\mu}$ 的第 i 个分量,σ_{ii} 为 $\boldsymbol{\Sigma}$ 的第 i 个对角元素.

推论 18.1.6

(1) 设 $\boldsymbol{x} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$，且 $\boldsymbol{\Sigma}>0$（正定），则

$\boldsymbol{y}=\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu}) \sim N(\boldsymbol{O},\boldsymbol{I})$，其中 $\boldsymbol{\Sigma}^{\frac{1}{2}}$ 为 $\boldsymbol{\Sigma}$ 的对称平方根，即 $\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^2=\boldsymbol{\Sigma}$，$\boldsymbol{\Sigma}^{\frac{1}{2}}$ 对称.

(2) 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu},\sigma^2\boldsymbol{I}_p)$，且 $\boldsymbol{G}$ 为 $q \times p$ 矩阵，满足 $\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}=\boldsymbol{I}_q$，则 $\boldsymbol{G}\boldsymbol{x} \sim N_q(\boldsymbol{G}\boldsymbol{\mu},\sigma^2\boldsymbol{I}_q)$.

定理 18.1.7 (1) 多元正态向量的 k 个互不相交的子向量是相互独立的，当且仅当它们是两两不相关的.

(2) 服从联合正态分布的两个向量是相互独立的，当且仅当它们的分量之间是两两独立的.

推论 18.1.8 (1) 若 $\boldsymbol{x} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$，则 $\boldsymbol{A}\boldsymbol{x}$ 和 $\boldsymbol{B}\boldsymbol{x}$ 相互独立，当且仅当 $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^{\mathrm{T}}=0$.

(2) 若 $\boldsymbol{x} \sim N(\boldsymbol{\mu},\sigma^2\boldsymbol{I})$，且 $\boldsymbol{G}$ 为任一行标准正交矩阵，即 $\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}=\boldsymbol{I}$，则 $\boldsymbol{G}\boldsymbol{x}$ 与 $(\boldsymbol{I}-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})\boldsymbol{x}$ 独立.

定理 18.1.9 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$\boldsymbol{x}^{\mathrm{T}}=(\boldsymbol{x}_1^{\mathrm{T}},\boldsymbol{x}_2^{\mathrm{T}})$，其中 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 分别为 $\boldsymbol{x}$ 的 r 元和 $q=p-r$ 元子向量. 记 $\boldsymbol{x}_{2\cdot 1}=\boldsymbol{x}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{x}_1$，$\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}$，$\boldsymbol{\mu}^{\mathrm{T}}=(\boldsymbol{\mu}_1^{\mathrm{T}},\boldsymbol{\mu}_2^{\mathrm{T}})$，则

$$\boldsymbol{x}_1 \sim N_r(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11}),$$
$$\boldsymbol{x}_{2\cdot 1} \sim N_{p-r}(\boldsymbol{\mu}_{2\cdot 1},\boldsymbol{\Sigma}_{22\cdot 1}),$$

其中 $\boldsymbol{\mu}_{2\cdot 1}=\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$，$\boldsymbol{\Sigma}_{22\cdot 1}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$，

且 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_{2\cdot 1}$ 相互独立.

定理 18.1.10 在定理 18.1.9 的假定和记号下有：给定 $\boldsymbol{x}_1$ 的条件下 $\boldsymbol{x}_2$ 的条件分布为

$$N_{p-r}(\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1),\boldsymbol{\Sigma}_{22\cdot 1}).$$

推论 18.1.11 设 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}} \sim N_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$，其中

$$\boldsymbol{\mu}=(\mu_1,\mu_2)^{\mathrm{T}},\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},$$

则在给定 x_1 的条件下 x_2 的条件分布为

$$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1),\sigma_2^2(1-\rho^2)\right).$$

定义 18.1.12 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 为取自总体 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的简单样本.记 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)^{\mathrm{T}}$,称为取自 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的数据矩阵.

定理 18.1.13 设 $\boldsymbol{X}$ 为取自 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的数据矩阵,记 $\bar{\boldsymbol{x}}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{1}$,其中 $\boldsymbol{1}=(1,\cdots,1)^{\mathrm{T}}$,则 $\bar{\boldsymbol{x}}\sim N_p\left(\boldsymbol{\mu},\frac{1}{n}\boldsymbol{\Sigma}\right)$. 又记 $\boldsymbol{S}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X}$,其中 $\boldsymbol{H}=\boldsymbol{I}-\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}$,则 $\bar{\boldsymbol{x}}$ 与 $\boldsymbol{S}$ 相互独立.

定理 18.1.14 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$,则 $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\sim\chi^2(p)$.

18.1.2 威沙特分布

在多元统计中,威沙特(Wishart)分布的作用相当于一元统计中的 χ^2 分布.

定义 18.1.15 设 $n\times p$ 矩阵 $\boldsymbol{X}$ 为取自 $N_p(\boldsymbol{0},\boldsymbol{\Sigma})$的数据矩阵,则 $\boldsymbol{M}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 的分布定义为具有自由度 n 和标度矩阵 $\boldsymbol{\Sigma}$ 的威沙特分布,记作 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$. 特别当 $\boldsymbol{\Sigma}=\boldsymbol{I}_p$ 时,称作标准威沙特分布.

推论 18.1.16 $W_1(\sigma^2,n)$与 $\sigma^2\chi^2(n)$有相同分布.

定理 18.1.17 设 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$且 $\boldsymbol{B}$ 为 $p\times q$ 常矩阵,则 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{B}\sim W_q(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{B},n)$.

推论 18.1.18 设 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$,则

(1) $\boldsymbol{M}$ 的任一对角子阵服从威沙特分布.

(2) $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Sigma}^{-\frac{1}{2}}\sim W_p(\boldsymbol{I},n)$.

(3) 若 $\boldsymbol{\Sigma}=\boldsymbol{I}$,且 $p\times q$ 阵 $\boldsymbol{B}$ 满足 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=\boldsymbol{I}_q$,则 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{B}\sim W_q(\boldsymbol{I},n)$.

定理 18.1.19 设 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$,且 $\boldsymbol{a}$ 为任一 p 元常向量,满

足 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma a}\neq 0$，则 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{Ma}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma a}\sim\chi^2(n)$.

推论 18.1.20 设 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$，$m_{ii}$为 $\boldsymbol{M}$ 的第 i 个对角元，则 $m_{ii}\sim\sigma_{ii}\chi^2(n)$，其中 σ_{ii} 为 $\boldsymbol{\Sigma}$ 的第 i 个对角元.

定理 18.1.21 设 $\boldsymbol{M}_i\sim W_p(\boldsymbol{\Sigma},n_i)$，$i=1,\cdots,k$，且 $\boldsymbol{M}_1,\cdots,\boldsymbol{M}_k$ 相互独立，则 $\sum_{i=1}^{k}\boldsymbol{M}_i\sim W_p(\boldsymbol{\Sigma},n)$，其中 $n=\sum_{i=1}^{k}n_i$.

定理 18.1.22 柯赫兰(Cochran)定理　设 $\boldsymbol{X}$ 为取自 $N_p(\boldsymbol{0},\boldsymbol{\Sigma})$的 $n\times p$ 数据阵，$\boldsymbol{C}$ 为 n 阶对称阵，则

(1) $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{CX}$ 的分布等价于 n 个独立 $W_p(\boldsymbol{\Sigma},1)$矩阵的加权和的分布，其中加权为 $\boldsymbol{C}$ 的特征值.

(2) $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{CX}$ 具有威沙特分布，当且仅当 $\boldsymbol{C}$ 为幂等阵，即 $\boldsymbol{C}^2=\boldsymbol{C}$. 此时 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{CX}\sim W_p(\boldsymbol{\Sigma},m)$，$m=\mathrm{rk}(\boldsymbol{C})=\mathrm{tr}(\boldsymbol{C})$.

(3) 设 $\boldsymbol{S}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{HX}$ 为样本方差阵，则

$$n\boldsymbol{S}\sim W_p(\boldsymbol{\Sigma},n-1).$$

定理 18.1.23 设 $\boldsymbol{X}$ 为取自 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的 $n\times p$ 数据阵，$\boldsymbol{C}_1,\cdots,\boldsymbol{C}_k$ 为 n 阶对称矩阵，则 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_1\boldsymbol{X},\cdots,\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_k\boldsymbol{X}$ 为相互独立的，当且仅当 $\boldsymbol{C}_r\boldsymbol{C}_s=0$，$r\neq s$.

以下考虑分块的威沙特矩阵. 先给出下列的记号.

$$\boldsymbol{M}=\begin{pmatrix}\boldsymbol{M}_{11} & \boldsymbol{M}_{12}\\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22}\end{pmatrix}\begin{matrix}r\\ p-r\end{matrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\begin{matrix}r\\ p-r\end{matrix},$$
$$\qquad\quad\ \ r\qquad p-r\qquad\qquad\qquad r\qquad p-r$$

$$\boldsymbol{M}_{22\cdot 1}=\boldsymbol{M}_{22}-\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12},\quad \boldsymbol{\Sigma}_{22\cdot 1}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.$$

定理 18.1.24 设 $\boldsymbol{M}\sim W_p(\boldsymbol{\Sigma},n)$，$n>r$，则

(1) $\boldsymbol{M}_{22\cdot 1}\sim W_{p-r}(\boldsymbol{\Sigma}_{22\cdot 1},n-r)$且独立于$(\boldsymbol{M}_{11},\boldsymbol{M}_{12})$.

(2) 若 $\boldsymbol{\Sigma}_{12}=0$，则 $\boldsymbol{M}_{22}-\boldsymbol{M}_{22\cdot 1}=\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}\sim W_{p-r}(\boldsymbol{\Sigma}_{22},r)$，且 $\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}$，$\boldsymbol{M}_{11}$及 $\boldsymbol{M}_{22\cdot 1}$相互独立.

记

$$M^{-1}=\begin{pmatrix}M^{11} & M^{12}\\ M^{21} & M^{22}\end{pmatrix},\Sigma^{-1}=\begin{pmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{pmatrix},$$

易见

$$M^{22}=M_{22\cdot 1}^{-1},\text{及 } \Sigma^{22}=\Sigma_{22\cdot 1}^{-1}.$$

推论 18.1.25 设 $M\sim W_p(\Sigma,n),n>r$,则

(1) $(M^{22})^{-1}\sim W_{p-r}((\Sigma^{22})^{-1},n-r)$且独立于$(M_{11}M_{12})$.

(2) 若 $\Sigma_{12}=0$,则 $M_{22}-(M^{22})^{-1}\sim W_{p-r}(\Sigma_{22},r)$且 $M_{22}-(M^{22})^{-1}$,M_{11}和$(M^{22})^{-1}$相互独立.

定理 18.1.26 设 $M\sim W_p(\Sigma,n)$且 $n>p$,则

(1) 对任意非零 p 元常向量,比值

$$a^{T}\Sigma^{-1}a/a^{T}M^{-1}a\sim\chi^2(n-p+1),$$

特别,设 m^{ii}和 σ^{ii}分别为 M^{-1}和 Σ^{-1}的第 i 个对角元,则 $\sigma^{ii}/m^{ii}\sim\chi^2(n-p+1)$.

(2) m^{ii}与 M 中除了第 i 个对角元 m_{ii}之外的所有元素独立.

定理 18.1.27 设 $M\sim W_p(\Sigma,n)$且 $n\geqslant p$,则$|M|$的分布等价于$|\Sigma|\cdot z_1\cdots,z_p$ 的分布,其中 $z_i\sim\chi^2(n-i+1),i=1,\cdots,p$,且 $z_1,\cdots,z_p$ 相互独立.

18.1.3 T^2 分布

T^2 分布是由霍特林(Hotelling)于 1931 年首先进行研究的,其作用相当于一元统计中的 t 分布.

定义 18.1.28 设 $\alpha=md^{T}M^{-1}d$,其中 $d\sim N_p(O,I)$,$M\sim W_p(I,m)$,且 d 与 M 相互独立,则称 α 的分布为具参数 p 和 m 的 T^2 分布,记作 $\alpha\sim T^2(p,m)$.

定理 18.1.29 设 x 与 M 为独立的各具分布 $N_p(\mu,\Sigma)$及 $W_p(\Sigma,m)$的随机变量,则

$$m(x-\mu)^{T}M^{-1}(x-\mu)\sim T^2(p,m).$$

推论 18.1.30 设 $\boldsymbol{X}$ 为取自 $N_p(\mu,\Sigma)$ 的 $n\times p$ 数据矩阵，$\bar{\boldsymbol{x}}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\mathbf{1}$ 为样本均值，$\boldsymbol{S}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X}$ 为样本方差矩阵，则

$$(n-1)(\bar{\boldsymbol{x}}-\mu)^{\mathrm{T}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\mu)\sim T^2(p,n-1).$$

上式左端的统计量常称为 T^2 统计量.

定理 18.1.31 $T^2(p,m)=\frac{mp}{m-p+1}F(p,m-p+1)$.

推论 18.1.32 设 $\boldsymbol{d}$ 与 $\boldsymbol{M}$ 服从定义 18.1.28 的条件，则

(1) $|\boldsymbol{M}|/|\boldsymbol{M}+\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}|\sim B\left(\frac{1}{2}(m-p+1),\frac{p}{2}\right)$：参数为 $\left(\frac{1}{2}(m-p+1),\frac{p}{2}\right)$ 的 β 分布.

(2) $\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}\left(1+\frac{1}{\boldsymbol{d}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{d}}\right)\sim\chi^2(m+1)$，且与 $\boldsymbol{d}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{d}$ 相互独立.

类似于一元统计的两样本 t 统计量，也有两样本的 T^2 统计量.

定理 18.1.33 设 $\boldsymbol{X}_i$ 为取自 $N_p(\mu_i,\Sigma)$ 的 $n_i\times p$ 数据阵，$i=1,2$. 若 $\mu_1=\mu_2$，则

$$(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2)^{\mathrm{T}}\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2)\sim\frac{n_1+n_2}{n_1n_2}T^2(p,n_1+n_2-2),$$

其中 $\bar{\boldsymbol{x}}_1,\bar{\boldsymbol{x}}_2$ 分别为两个样本的样本均值，$\boldsymbol{S}_p=(n_1\boldsymbol{S}_1+n_2\boldsymbol{S}_2)/(n_1+n_2-2)$，$\boldsymbol{S}_1,\boldsymbol{S}_2$ 分别为两样本的样本方差阵.

18.1.4 基于威沙特分布的统计量

在一元方差分析中常遇到基于独立 χ^2 分布随机变量比值的统计量，其相应的分布为 F 分布. 在多元统计中起同样作用的是 Λ 统计量和 Λ 分布.

定义 18.1.34 设 $\boldsymbol{A}\sim W_p(\boldsymbol{I},m)$，$\boldsymbol{B}\sim W_p(\boldsymbol{I},n)$ 相互独立且 $m\geqslant p$. 称

$$|\boldsymbol{A}|/|\boldsymbol{A}+\boldsymbol{B}|=|\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B}|^{-1}$$

服从具参数 p,m 和 n 的 **Λ 分布**，记作 $\Lambda(p,m,n)$.

定理 18.1.35 $\Lambda(p,m,n)$等价于 $\prod_{i=1}^{n} u_i$ 的分布，其中 $u_1,\cdots,u_n$ 相互独立，$u_i \sim B\left(\frac{1}{2}(m+i-p),\frac{p}{2}\right)$，$i=1,\cdots,n$.

定理 18.1.36 $\Lambda(p,m,n)=\Lambda(n,m+n-p,p)$. 根据 β 分布与 F 分布的关系(见 6.1.12)有

推论 18.1.37

(1) $(1-\Lambda(p,m,1))/\Lambda(p,m,1)=\dfrac{p}{m-p+1}F(p,m-p+1)$.

(2) $(1-\Lambda(1,m,n))/\Lambda(1,m,n)=\dfrac{n}{m}F(n,m)$.

(3) $(1-\sqrt{\Lambda(p,m,2)})/\sqrt{\Lambda(p,m,2)}$

$=\dfrac{p}{m-p+1}F(2p,2(m-p+1))$.

(4) $(1-\sqrt{\Lambda(2,m,n)})/\sqrt{\Lambda(2,m,n)}$

$=\dfrac{n}{m-1}F(2n,2(m-1))$.

定理 18.1.38 巴特莱特(Bartlett)定理　当 $m\to\infty$时，

$$-\{m-\frac{1}{2}(p-n+1)\}\log\Lambda(p,m,n)\xrightarrow{\mathscr{L}}\chi^2(np).$$

其中$\xrightarrow{\mathscr{L}}$表示依分布收敛.

定义 18.1.39 设 $\boldsymbol{A}\sim W_p(\boldsymbol{I},m)$，$m>p$，$\boldsymbol{B}\sim W_p(\boldsymbol{I},n)$，且 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 独立，则$(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}$ 的最大特征根 θ 称为**最大根统计量**(the largest root statistic)，其分布记作 $\theta(p,m,n)$.

推论 18.1.40 设 λ_1 为 $\boldsymbol{A}^{-1}\boldsymbol{B}$ 的最大特征根，则最大根统计量 $\theta=\lambda_1/(1+\lambda_1)$，因此 $0<\theta<1$.

推论 18.1.41

(1) $\theta(p,m,n)=\theta(n,m+n-p,p)$.

(2) $\dfrac{\theta(1,m,n)}{1-\theta(1,m,n)}=\dfrac{1-\Lambda(1,m,n)}{\Lambda(1,m,n)}=\dfrac{n}{m}F(n,m)$.

(3) $\frac{\theta(p,m,1)}{1-\theta(p,m,1)}=\frac{1-\Lambda(p,m,1)}{\Lambda(p,m,1)}$

$$=\frac{p}{m-p+1}F(p,m-p+1).$$

18.2 参数估计

18.2.1 多元正态分布的参数估计

设 $\boldsymbol{X}(n\times p)$ 为取自正态总体 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的数据阵，$\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)^{\mathrm{T}}$，记

$$\bar{\boldsymbol{x}}=\frac{1}{n}\sum_1^n \boldsymbol{x}_i,\quad \boldsymbol{S}=\frac{1}{n}\sum_1^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^{\mathrm{T}}. \tag{18.1}$$

$\bar{\boldsymbol{x}}$ 称为**样本均值**，$\boldsymbol{S}$ 称为**样本方差阵**.

$\boldsymbol{X}$ 的对数似然函数可表为如下形式：

$$l(\boldsymbol{\mu},\boldsymbol{\Sigma})=-\frac{n}{2}\log|2\pi\boldsymbol{\Sigma}|-\frac{n}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})$$

$$-\frac{n}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})(\bar{\boldsymbol{x}}-\boldsymbol{\mu})^{\mathrm{T}}). \tag{18.2}$$

由此可以推出 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的极大似然估计为

$$\hat{\boldsymbol{\mu}}=\bar{\boldsymbol{x}},\quad \hat{\boldsymbol{\Sigma}}=\boldsymbol{S}, \tag{18.3}$$

其中 $\bar{\boldsymbol{x}}$ 也是 $\boldsymbol{\mu}$ 的无偏估计，而 $\boldsymbol{S}$ 不是 $\boldsymbol{\Sigma}$ 的无偏估计. 但对 $\boldsymbol{S}$ 作一修正后便可得到 $\boldsymbol{\Sigma}$ 的无偏估计：

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{unbiase}}=\frac{n}{n-1}\boldsymbol{S}=\frac{1}{n-1}\sum_1^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^{\mathrm{T}}. \tag{18.4}$$

在某些特定的场合下对均值 $\boldsymbol{\mu}$ 和方差阵 $\boldsymbol{\Sigma}$ 会有一些约束，对均值 $\boldsymbol{\mu}$，我们假定有下述约束：

$$\boldsymbol{R}\boldsymbol{\mu}=\boldsymbol{r}, \tag{18.5}$$

其中 $\boldsymbol{R}$ 为 $q\times p$ 常数阵，$q<p$，$\boldsymbol{r}$ 为 q 维常向量.

定理 18.2.1 设 $\boldsymbol{X}(n\times p)$ 为取自 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的数据阵，$\boldsymbol{\mu}$ 满足约束条件(18.5)，则 $\boldsymbol{\mu}$ 的极大似然估计为

$$\hat{\mu} = \bar{x} - SR^{T}(RSR^{T})^{-1}(R\bar{x} - r).$$

Σ的极大似然估计为

$$\hat{\Sigma} = S + (\bar{x} - \hat{\mu})(\bar{x} - \hat{\mu})^{T}.$$

约束条件(18.5)的一个特例是

$$\mu = k\mu_0, \tag{18.6}$$

其中 μ_0 为给定的 p 维常向量，k 为未知参数. 这相当于取 R 为$(p-1)\times p$ 维阵，$\mathrm{rk}(R)=p-1$，$R\mu_0=0$，而 $r=0$.

推论 18.2.2 在约束条件(18.6)下，k 的极大似然估计为

$$\hat{k} = \mu_0^{T}S^{-1}\bar{x}/\mu_0^{T}S^{-1}\mu_0.$$

定理 18.2.1 中的 $\hat{\mu}$ 和推论 18.2.2 中的 $\hat{k}$ 都是无偏估计.

对方差阵，我们考虑下述约束条件

$$\Sigma = k\Sigma_0, \tag{18.7}$$

其中 Σ_0 为已知正定阵，k 为未知参数.

定理 18.2.3 设 $X(n\times p)$为取自 $N_p(\mu,\Sigma)$的数据阵，Σ 满足约束条件(18.7)，则 k 的极大似然估计为

$$\hat{k} = \mathrm{tr}(\Sigma_0^{-1}S)/p,$$

这里 $\hat{k}$ 不是无偏估计. 要得到无偏估计需乘系数$\frac{n-1}{n}$.

以上讨论的是样本取自同一总体时的情形. 下面考虑样本取自不同总体时的情形. 一般，我们假定样本 $x_1,\cdots,x_n$ 相互独立，$x_i\sim N_p(\mu_i,\Sigma_i)$，$i=1,\cdots,n$. $X=(x_1,\cdots,x_n)^{T}$ 为数据阵，如果对 μ_i 和 Σ_i 都不加任何约束条件，就得到退化情形，即 $\hat{\mu}_i=x_i$，Σ_i 估计不存在，$i=1,\cdots,n$. 因此我们要施加约束条件. 首先我们假定

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}, \tag{18.8}$$

$X_k(n_k\times p)$为取自 $N_p(\mu_k,\Sigma)$的数据阵，$\sum_{1}^{m} n_k = n$.

这条件相当于数据取自 m 个正态总体，每个有不同的均值和相同的方差阵.

定理 18.2.4 在条件(18.8)下，方差阵 $\boldsymbol{\Sigma}$ 的极大似然估计为

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{k=1}^{m}n_k\boldsymbol{S}_k,$$

其中 $\boldsymbol{S}_k$ 为 $\boldsymbol{X}_k$ 的样本方差阵.

其次考虑同均值，不同方差阵的情形. 设 $\boldsymbol{X}$ 如式(18.8)中所示.

$$\boldsymbol{X}_k(n_k\times p)\text{为取自 }N_p(\boldsymbol{\mu},\boldsymbol{\Sigma}_k)\text{的数据阵},$$

$$k=1,\cdots,m. \tag{18.9}$$

定理 18.2.5 在约束条件(18.9)下，$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}_k$，$k=1,\cdots,m$ 的极大似然估计满足方程

$$\hat{\boldsymbol{\mu}}=\left(\sum_{k=1}^{m}n_k\ \hat{\boldsymbol{\Sigma}}_k\right)^{-1}\left(\sum_{k=1}^{m}n_k\ \hat{\boldsymbol{\Sigma}}_k^{-1}\overline{\boldsymbol{x}}_k\right),$$

$$\hat{\boldsymbol{\Sigma}}_k=\boldsymbol{S}_k+(\overline{\boldsymbol{x}}_k-\hat{\boldsymbol{\mu}})(\overline{\boldsymbol{x}}_k-\hat{\boldsymbol{\mu}})^{\mathrm{T}},k=1,\cdots,n.$$

其中 $\overline{\boldsymbol{x}}_k$ 和 $\boldsymbol{S}_k$ 分别为第 k 个子数据阵 $\boldsymbol{X}_k$ 的样本均值和样本方差阵，$k=1,\cdots,n$.

解上述联立方程可用迭代法：先用 $\boldsymbol{S}_k$ 作 $\hat{\boldsymbol{\Sigma}}_k$ 的初值，由此解得 $\hat{\boldsymbol{\mu}}$，再用 $\hat{\boldsymbol{\mu}}$ 代入第二组方程求得新的 $\hat{\boldsymbol{\Sigma}}_k$，如此反复，直至所求解不再变化为止.

最后我们考虑下述约束条件

$$\begin{aligned}&\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)^{\mathrm{T}},\boldsymbol{x}_i\sim N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}),\\&\boldsymbol{\mu}_i=\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{c}_i,\quad i=1,\cdots,n,\end{aligned} \tag{18.10}$$

其中 $\boldsymbol{\Theta}$ 为 $q\times p$ 未知参数阵，$\boldsymbol{c}_i$ 为 q 维已知常向量.

定理 18.2.6 记 $\boldsymbol{C}=(\boldsymbol{c}_1,\cdots,\boldsymbol{c}_n)^{\mathrm{T}}$，则 $\boldsymbol{\Theta}$ 的极大似然估计为

$$\hat{\boldsymbol{\Theta}}=(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{X}.$$

18.2.2 稳健估计

多元分布均值向量与方差阵的稳健估计在理论上比一元分布场合更为复杂. 本节只推荐几种较为实用的方法.

剔除异常值(outlier)的一种方法是从样品 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 中剔除掉 $\alpha 100\%$ 的样品. 其具体作法是先给定一个小的正数 $\alpha(=0.05$ 或 $0.01)$, 对第 j 个变量的样本值 $x_{kj}, k=1,\cdots,n$, 剔除最大的 $\frac{\alpha}{2p}n$ 个, 剔除最小的 $\frac{\alpha}{2p}n$ 个. 如果对某个变量 j, 样品值 x_{kj} 是在被剔除之列, 则整个样品 $\boldsymbol{x}_k$ 被剔除. 这样全部 n 个样品被剔除的个数不超过 αn. 余下的不少于 $(1-\alpha)n$ 个样品可用来计算其样本均值, 加权为 $1/(1-\alpha)n$, 作为均值向量的稳健估计.

另外一种值得推荐的求均值向量的稳健估计的方法是求 $\tilde{\mu}$ 满足下述方程.

$$\sum_1^n \|\boldsymbol{x}_k - \tilde{\boldsymbol{\mu}}\|^\gamma = \min_{\mu} \sum_1^n \|\boldsymbol{x}_k - \boldsymbol{\mu}\|^\gamma, \tag{18.11}$$

其中 $\|\boldsymbol{x}_k-\boldsymbol{\mu}\|$ 为欧氏模, $\left(\|\boldsymbol{x}_k-\boldsymbol{\mu}\| = \left[\sum_{i=1}^p (x_{ki}-\mu_i)^2\right]^{\frac{1}{2}}\right)$ $\gamma \in [1,2]$. 当 $\gamma<2$ 时, 异常值的影响被消弱(见 14.36). 但另一方面, 当 γ 较小(接近 1)时, 估计的精度也不够好. 一个折中的建议是令 $\gamma=3/2$.

下面介绍一种比较实用的样本方差阵的稳健估计方法.

第 1 步. 给出样本均值 μ 的一个稳健估计 $\tilde{\mu}$(可以采用上述两种方法之一). 将 n 个样品按与 $\tilde{\mu}$ 的欧氏距离从小到大进行排序.

第 2 步. 从样品中剔除 αn 个序号最大的, α 为事先给定的小的正数. 余下的样品集合记为 R, 计算 p 阶方阵

$$\boldsymbol{A} = \sum_{x_k \in R} (\boldsymbol{x}_k - \tilde{\boldsymbol{\mu}})(\boldsymbol{x}_k - \tilde{\boldsymbol{\mu}})^{\mathrm{T}}, \tag{18.12}$$

当 $(1-\alpha)n>p$ 时, 通常可保证 $\boldsymbol{A}$ 为正定阵.

第 3 步. 计算 $\|x_k-\tilde{\mu}\|_M^2=(x_k-\tilde{\mu})^T A^{-1}(x_k-\tilde{\mu}), k=1,\cdots,n.$ 按 $\|x_k-\tilde{\mu}\|_M$ 的值从小到大对样品进行重排序.

第 4 步. 对事先给定的小的正数 β,按第三步的重新排序剔除掉 βn 个序号最大的样品,余下的 $(1-\beta)n$ 个样品的集合记为 T. 方差阵 $\boldsymbol{\Sigma}$ 的稳健估计由下式给出.

$$\tilde{\tilde{\boldsymbol{\Sigma}}}=\frac{1}{n(1-\beta)}\sum_{x_k\in T}(x_k-\tilde{\mu})(x_k-\tilde{\mu})^T. \tag{18.13}$$

采用剔除异常值的方法构造均值和方差阵的稳健估计时,都要求 $(1-\alpha)n$ (或 $(1-\beta)n$) 大于 p. 从实用的观点来看至少要大于 $2p$,才能使所得估计有效.

18.3 假设检验

多元假设检验比较一元假设检验问题增加了许多复杂性. 首先,由于多元总体涉及的参数较多,因此可能设置的假设也更多,更复杂. 在实际问题中,要根据问题的背景来设置恰当的假设是一项技巧性较强的工作. 其次,当给定了假设之后,选择一个好的检验又是一项困难的任务. 在一元统计的假设检验理论中,可以根据一定的准则,在给定的检验类中寻找某个最优检验;而在多元假设检验问题中,由于上述的复杂性,是几乎不可能的. 本节将重点介绍一种在大多数情况下都可行的检验方法——似然比检验.

18.3.1 似然比检验

定义 18.3.1 设样本 $X=(x_1,\cdots,x_n)^T$ 有似然函数 $L(X,\theta)$,其中 θ 为未知参数向量,设

$$H_0: \theta\in\Omega_0, \qquad H_1: \theta\in\Omega_1$$

分别为原假设和备择假设,则定义**似然比统计量**(likelihood ratio statistic)为

$$\lambda = L_0/L_1,$$

其中，$L_0=\max\limits_{\theta\in\Omega_0}L(\boldsymbol{X},\theta)$，$L_1=\max\limits_{\theta\in\Omega_0\cup\Omega_1}L(\boldsymbol{X},\theta)$.

定义 18.3.2 对给定的小的正数 $\alpha\in(0,1)$，假设检验问题 H_0 对 H_1 的似然比检验(likelihood ratio test)有拒绝域

$$W = \{\boldsymbol{X}:\lambda(\boldsymbol{X}) < c\},$$

其中 c 由下式确定：

$$\sup_{\theta\in\Omega_0}P_\theta(\boldsymbol{X}\in W\} = \alpha.$$

似然比统计量具有下述的重要性质.

定理 18.3.3 设参数空间 $\Omega=\Omega_0\cup\Omega_1$ 为 R^q 中的一个区域，Ω_0 为 Ω 的一个 r 维子区域，$r<q$，则在一定正则条件下，对$\forall\ \theta\in\Omega_0$，$-2\log\lambda$，当 $n\to\infty$时有渐近分布 $\chi^2(q-r)$.

似然比检验有两个显著的优点. 首先，由于似然比统计量是由似然函数极值之比构成的. 因此，在各种不同参数的不同形式的假设检验问题中给出了一个统一的寻求检验统计量的思路. 其次，由定理 18.3.3 知，在相当广泛的分布类中，$-2\log\lambda$ 有渐近 χ^2 分布，因此在大样本的前提下，似然比检验的拒绝域比较容易确定.

下面的小节将着重讨论：单正态总体的均值检验，多正态总体的均值检验，正态总体的方差阵的检验及均值的联合置信区间的问题.

18.3.2 单正态总体的均值检验

设 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)^{\mathrm{T}}$ 为取自正态总体 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$的独立同分布样本矩阵.

方法 18.3.4 当 $\boldsymbol{\Sigma}$ 已知时，检验假设

$$H_0:\mu = \mu_0, H_1:\mu \neq \mu_0.$$

记 λ 为似然比统计量，有

$$-2\log\lambda = n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0).$$

由 $\sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\sim N_p(0,\boldsymbol{\Sigma})$，据定理 18.1.14 知：

$$-2\log\lambda \sim \chi^2(p),$$

因此对给定的水平 α，拒绝域为

$$W=\{-2\log\lambda > c\},$$

其中 c 为 $\chi^2(p)$的上 α 分位点.

方法 18.3.5 当 $\boldsymbol{\Sigma}$ 未知时，检验假设

$$H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0,H_1:\boldsymbol{\mu}\neq\boldsymbol{\mu}_0.$$

此时有

$$-2\log\lambda = n\log\left(1+\frac{1}{n-1}T^2\right),$$

其中

$$T^2=(n-1)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\sim T^2(p,n-1),$$

称为霍特林单样本 T^2 统计量. 因此，水平 α 的拒绝域为(注意到 $-2\log\lambda$是 T^2 的严格单调增函数)：

$$W=\{T^2>c\},$$

其中 c 为 $T^2(p,n-1)$的上 α 分位点. 当 T^2 分布表不易找到时，可利用 T^2 分布与 F 分布的关系(见定理 18.1.31)去查找 F 分布表来确定 c.

下面进一步介绍关于均值 $\boldsymbol{\mu}$ 的复杂假设的检验问题. 在关于 $\boldsymbol{\mu}$ 的检验中常会检验 $\boldsymbol{\mu}$ 的某个(或某些)线性组合是否等于某个(或某些)给定的值；或比较其中若干个分量是否相等；或考虑 $\boldsymbol{\mu}$ 是否与某个给定的向量成比例. 上述假设问题可统一地表示为如下的检验假设问题

$$H_0:\ \boldsymbol{R\mu}=\boldsymbol{r}, \tag{18.14}$$

其中 $\boldsymbol{R}$ 为$(q\times p)$常数阵，$rk(\boldsymbol{R})=q$，$\boldsymbol{r}$ 为 q 维常向量.

方法 18.3.6 当 $\boldsymbol{\Sigma}$ 已知时，检验假设(18.14)，此时，似然比统计量有形式：

$$-2\log\lambda = n(\boldsymbol{R}\bar{\boldsymbol{x}}-\boldsymbol{r})^{\mathrm{T}}(\boldsymbol{R\Sigma R}^{\mathrm{T}})^{-1}(\boldsymbol{R}\bar{\boldsymbol{x}}-\boldsymbol{r})\sim\chi^2(q).$$

由上式便可确定拒绝域.

方法 18.3.7 当 $\boldsymbol{\Sigma}$ 已知时,检验假设

H_0: $\boldsymbol{\mu}=k\boldsymbol{\mu}_0$,对某个 k 成立,

其中 $\boldsymbol{\mu}_0$ 为常向量.这是式(18.14)的一个特例.即相当取 $\boldsymbol{R}$ 为$(p-1)\times p$ 阵,$\boldsymbol{R}$ 的各行都与 $\boldsymbol{\mu}_0$ 正交,而取 $\boldsymbol{r}=0$.此时可推出

$$-2\log\lambda=n\bar{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{\Sigma}-(\boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0)^{-1}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^{\mathrm{T}}\}\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}\sim\chi^2(p-1).$$

方法 18.3.8 当 $\boldsymbol{\Sigma}$ 未知时,检验假设(18.14)

此时有

$$-2\log\lambda=n\log(1+\boldsymbol{d}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{d}),$$

其中 $\boldsymbol{d}=\boldsymbol{S}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^{\mathrm{T}})^{-1}(\boldsymbol{R}\bar{\boldsymbol{x}}-\boldsymbol{r})$,

$$\boldsymbol{d}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{d}=(\boldsymbol{R}\bar{\boldsymbol{x}}-\boldsymbol{r})^{\mathrm{T}}(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^{\mathrm{T}})^{-1}(\boldsymbol{R}\bar{\boldsymbol{x}}-\boldsymbol{r})\sim\frac{1}{n-1}T^2(q,n-1).$$

由于 $-2\log\lambda$ 是 $\boldsymbol{d}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{d}$ 的严格单调增函数,容易用 $\boldsymbol{d}^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{d}$ 来确定拒绝域.

方法 18.3.9 当 $\boldsymbol{\Sigma}$ 未知时,检验 18.3.7 中的假设.

此时有

$$-2\log\lambda=n\log\left(1+\frac{1}{n-1}T^2\right),$$

其中

$$T^2=(n-1)\bar{\boldsymbol{x}}\boldsymbol{S}^{-1}\{\boldsymbol{S}-(\boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{S}^{-1}\boldsymbol{\mu}_0)^{-1}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^{\mathrm{T}}\}\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}$$
$$\sim T^2(p-1)(n-1).$$

因此拒绝域为 $W=\{T^2>c\}$.

18.3.3 多正态总体的均值检验

设 $\boldsymbol{X}_i(n_i\times p)$ 为取自总体 $N_p(\boldsymbol{\mu}_i\boldsymbol{\Sigma}_i)$ 的数据阵,$i=1,\cdots,k$.

方法 18.3.10 在 $\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_k$ 的条件下,检验假设

$$H_0:\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_k.$$

为给出似然比统计量,先规定如下记号,记 $\bar{\boldsymbol{x}}_i$ 和 $\boldsymbol{S}_i$ 为 $\boldsymbol{X}_i$ 的样

本均值和样本方差阵，$i=1,\cdots,k$. 又记 k 个样本的合样本为

$$X=(X_1^{\mathrm{T}},\cdots,X_k^{\mathrm{T}})^{\mathrm{T}}.$$

合样本的样本均值为 $\bar{x}$，样本方差阵为 S. 则似然比统计量有形式

$$\lambda=\left\{\frac{|W|}{|nS|}\right\}^{\frac{n}{2}}=|T^{-1}W|^{\frac{n}{2}},$$

其中 $W=\sum_{i=1}^{k}n_iS_i$，$T=nS$. T 称为全离差阵，W 称为组内离差阵. 记

$$B=T-W=\sum_{i=1}^{k}n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^{\mathrm{T}},$$

称为组间离差阵.

定理 18.3.11 $|T^{-1}B|=|(B+W)^{-1}W|=|I+W^{-1}B|^{-1}$服从 $\Lambda(p,n-k,k-1)$分布，当 $n\geqslant p+k$.

由上述 λ 的表示及定理便可给出检验问题 18.3.10 的似然比检验的拒绝域.

方法 18.3.12 两样本霍特林 T^2 检验 在检验假设 18.3.10 时，当 $k=2$，似然比统计量

$$|I+W^{-1}B|^{-1}=\left\{1+\frac{n_1n_2}{n_1+n_2}(\bar{x}_1-\bar{x}_2)^{\mathrm{T}}W^{-1}(\bar{x}_1-\bar{x}_2)\right\}^{-1}$$

$$=\{1+T^2\}^{-1},$$

其中 $\mathrm{T}^2\sim\mathrm{T}^2(p,n_1+n_2-2)$为两样本 T^2 统计量.

在此特殊场合下，似然比检验转化为 T^2 检验.

方法 18.3.13 当 $\Sigma_1,\cdots,\Sigma_k$ 不相等且未知时，检验假设

$$H_0:\ \mu_1=\cdots=\mu_k.$$

这是一元统计中贝伦斯-费希尔(Behrens-Fisher)问题的推广. 似然比统计量有形式

$$-2\log\lambda=\sum_{i=1}^{k}\left\{n_i\log\left(\frac{|\hat{\Sigma}_i|}{|S_i|}\right)+\mathrm{t_r}(\hat{\Sigma}_i^{-1}S_i)\right.$$

$$+ n_i(\bar{\boldsymbol{x}}_i - \hat{\mu})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_i^{-1}(\bar{\boldsymbol{x}}_i - \hat{\mu})\} - \frac{1}{2}np.$$

其中 $\hat{\mu}$ 和 $\hat{\boldsymbol{\Sigma}}_1, \cdots, \hat{\boldsymbol{\Sigma}}_k$ 满足下列联立方程：

$$\hat{\mu} = \left(\sum_{i=1}^{k} n_i \hat{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{k} n_i \hat{\boldsymbol{\Sigma}}_i^{-1}\bar{\boldsymbol{x}}_i\right), \tag{18.15}$$

$$\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i + (\bar{\boldsymbol{x}}_i - \hat{\mu})(\bar{\boldsymbol{x}}_i - \hat{\mu})^{\mathrm{T}}, i = 1, \cdots, k.$$

解上述方程组可用迭代法：先用 $\boldsymbol{S}_i$ 作 $\hat{\boldsymbol{\Sigma}}_i$ 的初值，由此解得 $\hat{\mu}$，再求得新的 $\hat{\boldsymbol{\Sigma}}_i$，如此多次迭代，直至所求的解收敛为止.

18.3.4 方差矩阵的检验

设数据阵 $\boldsymbol{X}(n \times p)$ 取自总体 $N_p(\mu, \boldsymbol{\Sigma})$.

方法 18.3.14 在 μ 未知的条件下，检验假设

$$H_0: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0.$$

似然比有形式

$$-2\log\lambda = n\mathrm{t_r}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}) - n\log|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}| - np.$$

$-2\log\lambda$ 的分布是非常复杂的，但由定理 18.3.3 知，$-2\log\lambda$ 有渐近分布 $\chi^2\left(\frac{p}{2}(p+1)\right)$，因此当 n 较大时，可用此分布近似计算此检验的拒绝域. 为了更精确起见，我们可用 $(1-k)\chi^2\left(\frac{1}{2}p(p+1)\right)$ 来作 $-2\log\lambda$ 的渐近分布，其中

$$k = \left(2p + 1 - \frac{2}{p+1}\right)\Big/6(n-1). \tag{18.16}$$

当 n 较小时，可查 $L(m, v)$ 表.

方法 18.3.15 球性检验 考虑如下的检验问题：

$H_0: \boldsymbol{\Sigma} = k\boldsymbol{I}$，其中 k 为某正数.

这假设相当于说各分量相互独立且有相同方差，这时似然比有形式

$$-2\log\lambda = np(\mathrm{t_r}\boldsymbol{S} - \log|\boldsymbol{S}| - 1),$$

其极限分布为 $\chi^2\left(\frac{1}{2}(p-1)(p+2)\right)$.

方法 18.3.16 检验假设 $H_0: \boldsymbol{\Sigma}_{12}=0$.

其中 $\boldsymbol{\Sigma}_{12}(q\times(p-q))$为 $\boldsymbol{\Sigma}$ 的右上角子阵.

记
$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\begin{matrix}q\\ (p-q)\end{matrix},$$
$$\qquad\qquad\quad q \qquad (p-q)$$

显然,此假设是要检验 $\boldsymbol{x}$ 的前 q 个分量与后 $p-q$ 个分量的独立性问题.此时似然比有形式

$$\lambda = |\boldsymbol{I} - \boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}|^{\frac{n}{2}},$$

其中

$$\boldsymbol{S} = \begin{bmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{bmatrix},$$

$$\lambda^{\frac{2}{n}} \sim \Lambda(p-q, n-1-q, q).$$

其水平 α 的拒绝域为

$$W = \{\lambda^{\frac{2}{n}} < c\}.$$

c 可通过查 Λ 分布表求得.

方法 18.3.17 检验假设 H_0:$\boldsymbol{\Sigma}$ 为对角阵.

这个假设是要检验 $\boldsymbol{x}$ 的各分量相互独立(不相关).此时似然比有形式

$$-2\log\lambda = -n\log|\boldsymbol{R}|,$$

其中 $\boldsymbol{R}$ 为样本相关阵,其渐近分布为

$$-\left(n-\frac{1}{6}(2p+11)\right)\log|\boldsymbol{R}| \sim \chi^2\left(\frac{1}{2}p(p-1)\right).$$

以上讨论的都是单样本的方差阵的检验问题.下面讨论多样本方差阵的检验问题.我们以下采用 18.3.3 小节中的记号.

方法 18.3.18 当 $\mu_1,\cdots,\mu_k$ 未知时,检验假设

$$H_0: \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k.$$

这检验称为方差阵的齐次性(homogeneity)检验.似然比有形式

$$-2\log\lambda = n\log|\boldsymbol{S}| - \sum n_i\log|\boldsymbol{S}_i| = \sum n_i\log|\boldsymbol{S}_i^{-1}\boldsymbol{S}|,$$

其中 n, $\boldsymbol{S}_i$ 和 $\boldsymbol{S}$ 的定义见 18.3.3 小节.

$-2\log\lambda$ 有渐近分布 $\chi^2\left(\frac{1}{2}p(p+1)(k-1)\right)$.

方法 18.3.19 检验假设

$$H_0:\ \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k \text{ 且 } \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k.$$

这检验称为完全齐次性检验.似然比有形式

$$-2\log\lambda = n\log\left|\frac{1}{n}\boldsymbol{W}\right| - \sum_{i=1}^{n} n_i\log|\boldsymbol{S}_i|,$$

其中 $\boldsymbol{W} = \sum_{i=1}^{k} n_i\boldsymbol{S}_i$. $-2\log\lambda$ 的渐近分布为

$$\chi^2\left(\frac{1}{2}p(k-1)(p+3)\right).$$

18.3.5 并交检验

除了似然比检验外,另一类用得较多的检验方法称为**并交检验**(union intersection test),又称为**人工一元化检验**(artificial univariate test).其基本思想如下.

考虑一个随机向量 x,其分布为 F_θ,其中 θ 为未知参数向量.假定要检验假设 H_0: $\boldsymbol{\theta}\in\Theta_0$.任取常向量 $\boldsymbol{a}$,令 $y_a=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$.设在 H_0 下,y_a 作为一元变量,其分布参数 $\delta(\boldsymbol{\theta},\boldsymbol{a})$(一元参数)应满足 H_{0a}: $\delta\in\Delta_{0a}$.因此当给定 $\boldsymbol{a}$ 时,在多元场合检验 H_0 的问题转化成在一元场合检验 H_{0a} 的问题.且 H_0 成立当且仅当对任给的 $\boldsymbol{a}$, H_{0a} 成立,即

$$H_0 = \bigcap_a H_{0a}. \tag{18.17}$$

为检验 H_{0a} 成立与否,假定我们由 y_a 构造了统计量 z_a,对 H_{0a}

的拒绝域为 $W_a=\{z_a:z_a>c\}$. 由(18.17)知,拒绝任一个 H_{0a} 都意味着拒绝 H_0,因此,我们可得到 H_0 的拒绝域为

$$W=\bigcup_a W_a. \tag{18.18}$$

由(18.17),(18.18)两式便给出这种检验方法的名称:并交检验.

限于篇幅,我们只给出并交检验方法的两个例子.

例 18.3.20 设 X 为取自 $N_p(\mu,\boldsymbol{\Sigma})$的$(n\times p)$数据阵,$\boldsymbol{\Sigma}$ 未知用并交法检验 $H_0:\mu=\mu_0$.

、 记 $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{a}^{\mathrm{T}}$,则 $\boldsymbol{y}$ 的元素为取自总体 $N(\boldsymbol{a}^{\mathrm{T}}\mu,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{a})$的独立同分布样本,相应于 H_0 的一元化假设为 $H_{0a}:\boldsymbol{a}^{\mathrm{T}}\mu=\boldsymbol{a}^{\mathrm{T}}\mu_0$. 熟知,对 H_{0a},通常的 t 检验为令

$$t_a=(\bar{y}-\mu_y)/\sqrt{S_y^2/(n-1)}. \tag{18.19}$$

其中 $\bar{y}$ 为 $\boldsymbol{y}$ 的样本均值,S_y^2 为 y 的样本方差,$\mu_y=\boldsymbol{a}^{\mathrm{T}}\mu$. 因此,$H_{0a}$的拒绝域为

$$W_a=\{t_a^2>c\}. \tag{18.20}$$

注意,因 t_a^2 的分布与 $\boldsymbol{a}$ 无关,因此在上式中 c 与 $\boldsymbol{a}$ 无关,我们有

$$t_a^2=(n-1)\frac{\boldsymbol{a}^{\mathrm{T}}(\bar{\boldsymbol{x}}-\mu_0)(\bar{\boldsymbol{x}}-\mu_0)^{\mathrm{T}}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{a}}. \tag{18.21}$$

由上式易推出

$$\max_a t_a^2=(n-1)(\bar{\boldsymbol{x}}-\mu_0)^{\mathrm{T}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\mu_0)=T^2. \tag{18.22}$$

其中 T^2 为熟知的霍特林(Hotelling)单样本 T^2 统计量. 因此原假设 H_0 的拒绝域为

$$W=\bigcup_a W_a=\{\max_a t_a>c\}=\{T^2>c\}. \tag{18.23}$$

由(18.22)可见,在这一场合并交检验与似然比检验是一致的.

例 18.3.21 设 $\boldsymbol{X}$ 为取自 $N_p(\mu,\boldsymbol{\Sigma})$的$(n\times p)$数据阵,$\mu$ 未知,

用并交法检验假设 $H_0:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$.

令 $\boldsymbol{y}$ 如例 18.3.20 中定义,其分布如所述 H_{0a}: $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{a}$. 一元统计检验的拒绝域为

$$W_a = \{U_a < c_1, \text{或 } U_a > c_2\}, \tag{18.24}$$

其中

$$U_a = n\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{a}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{a}. \tag{18.25}$$

注意 U_a 的分布与 $\boldsymbol{a}$ 无关,因此 c_1,c_2 与 $\boldsymbol{a}$ 无关. 易算出

$$\max_a U_a = n\lambda_1, \min_a U_a = n\lambda_p. \tag{18.26}$$

其中 λ_1 为 $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}$ 的最大特征根,而 λ_p 为 $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}$ 的最小特征根. 因此 H_0 的拒绝域为

$$W = \bigcup_a W_a = \{\lambda_p < c_1, \text{或 } \lambda_1 > c_2\}. \tag{18.27}$$

在这场合下,并交检验与似然比检验不同.

18.3.6 联合置信区间

并交检验法的另一个重要应用是可用来构造参数的联合置信区间.

设 $\boldsymbol{X}$ 为取自 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的数据矩阵,我们要对 p 维参数向量 $\boldsymbol{\mu}$ 给出一个水平为 $1-\alpha$ 的联合区间. 设 $\boldsymbol{\mu}=(\mu_1,\cdots,\mu_p)^{\mathrm{T}}$,要给出区间 $[a_i,b_i]$,$i=1,\cdots,p$,使得

$$P(\mu_i \in [a_i,b_i], i = 1,\cdots,p) \geqslant 1 - \alpha.$$

更一般地说,对 $\boldsymbol{\mu}$ 的任一线性组合 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu}$,要给出区间 Δ_a,使得

$$P(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} \in \Delta_a, \forall\, \boldsymbol{a}) = 1 - \alpha.$$

由(18.19)知

$$t_a = \boldsymbol{a}^{\mathrm{T}}(\overline{\boldsymbol{x}} - \boldsymbol{\mu})/[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{a}/(n-1)]^{\frac{1}{2}}.$$

且由(18.21)知

$$t_a^2 \leqslant T^2.$$

因此若取适当的 c 使得

$$P(T^2 > c) = \alpha.$$

于是有

$$P(t_a^2 \leqslant c, 一切\ \boldsymbol{a}) = 1 - \alpha.$$

因此可解得 μ 的联合区间为

$$P(\boldsymbol{a}^{\mathrm{T}}\mu \in [\boldsymbol{a}^{\mathrm{T}}\bar{\boldsymbol{x}} - b, \boldsymbol{a}^{\mathrm{T}}\bar{\boldsymbol{x}} + b], 一切\ \boldsymbol{a}) = 1 - \alpha,$$

其中，对固定的 $\boldsymbol{a}$ 有

$$b = c[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{a}/(n-1)]^{\frac{1}{2}}.$$

18.4 多元回归分析

18.4.1 模型及参数估计

多元回归分析(multivariate regression analysis)是一元回归分析在多元场合下的推广. 在一元统计中也有所谓多元回归(multiple regression)(见第 15 章)，即指多个自变量(又称预报因子)对 1 个因变量(又称预报量)的回归，确切地说应翻译成多重回归. 本节所说的多元回归是指多个自变量对多个因变量的回归.

设有 q 个自变量 $x_1, \cdots, x_q$，p 个因变量 $y_1, \cdots, y_p$. 理论上可以假设 y_i 与 x_i 之间有下述关系：

$$y_i = \mu_i + \beta_{1i}x_1 + \beta_{2i}x_2 + \cdots + \beta_{qi}x_q + \varepsilon_i, i = 1, \cdots, p.$$

又设当给定 $x_1, \cdots, x_q$ 的值时有

$$E(\varepsilon_i) = 0, \qquad \mathrm{cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij}, \qquad i, j = 1, \cdots, p.$$

上述二式可用矩阵和向量的形式表示. 记

$$\boldsymbol{y} = (y_1, \cdots, y_p), \boldsymbol{x} = (x_1, \cdots, x_q),$$

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p), \boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p),$$

$$\boldsymbol{B} = \begin{bmatrix} \beta_{11}\beta_{12} & \cdots & \beta_{1p} \\ \beta_{21}\beta_{22} & \cdots & \beta_{2p} \\ \cdots & \cdots & \cdots \\ \beta_{q1}\beta_{q2} & \cdots & \beta_{qp} \end{bmatrix}.$$

于是，多元回归模型可表为

模型 18.4.1 $\begin{cases} y=(1,x)\begin{pmatrix}\mu\\ B\end{pmatrix}+\varepsilon=\underline{x}\,\underline{B}+\varepsilon, \\ \text{给定 } x \text{ 时}, E(\varepsilon)=0, \operatorname{cov}(\varepsilon^{T})=\Sigma. \end{cases}$

其中$\underline{x}=(1,x)$，$\Sigma=(\sigma_{ij})$，$\underline{B}=\begin{pmatrix}\mu\\ B\end{pmatrix}$. 参数矩阵$\underline{B}$ 和方差阵 Σ 是待估计的.

注 1 在 18.4.1 给出的多元回归的理论模型中自变量 x 可以看成是可精确观测的，也可以看成是随机变量. 在后一场合总假定 x 与 ε 独立，因此这两种观点是等价的.

注 2 在 18.4.1 中，注意 y,x,μ,ε 均是行向量，这是本节中为简化表达式所采取的一种特殊处理方法，希望读者注意.

设对 x 和 y 经过 n 次观测，得到 n 组观测值 $x_{(i)}$，$y_{(i)}$，$i=1,\cdots,n$，其中 $x_{(i)}$和 $y_{(i)}$分别是 q 维和 p 维的行向量. 记

$$\underset{n\times p}{Y}=\begin{pmatrix}y_{(1)}\\ \vdots\\ y_{(n)}\end{pmatrix},\quad \underset{n\times q}{X}=\begin{pmatrix}x_{(1)}\\ \vdots\\ x_{(n)}\end{pmatrix}.$$

于是我们可由 18.4.1 所给出的理论模型得出多元回归观测数据的具体模型为

模型 18.4.2 $\begin{cases} Y=(1_n,X)\begin{pmatrix}\mu\\ B\end{pmatrix}+E=\underline{X}\ \underline{B}+E, \\ \text{给定 } X \text{ 时}, E \text{ 的各行不相关，每行有均值零，方差阵 } \Sigma. \end{cases}$

其中 1_n 为各元素为 1 的列向量. 多元回归的一个最主要的任务就是从具体模型 18.4.2 出发，给出参数阵$\underline{B}$ 的估计$\hat{\underline{B}}^{T}=(\hat{\mu},\hat{B})^{T}$，并利用下述的预报方程

$$\hat{y}=(1,x)\begin{pmatrix}\hat{\mu}\\ \hat{B}\end{pmatrix}=\hat{\mu}+x\hat{B} \tag{18.28}$$

去预测 y 在 x 的新观测值下的预报值.

下面给出一个多元回归的例子.

例 18.4.3 趋势面分析 设 $\boldsymbol{y}$ 表示 p 种矿物的含量. 在某一地区的 n 个地点测量 $\boldsymbol{y}$ 的值,得到观测数据 $\boldsymbol{y}_{(1)},\cdots,\boldsymbol{y}_{(n)}$. 这 n 个地点的地理座标分别记录为 (u_i,v_i), $i=1,\cdots,n$. 理论上可以假设 $\boldsymbol{y}$ 是地理座标 (u,v) 的连续函数,因而可以用一个关于 u,v 的多项式来逼近. 设用二次多项式来逼近,于是在 18.4.2 中我们可取

$$\underline{\boldsymbol{X}}=\begin{pmatrix}1 & u_1 & v_1 & u_1^2 & u_1v_1 & v_1^2\\ 1 & u_2 & v_2 & u_2^2 & u_2v_2 & v_2^2\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & u_n & v_n & u_n^2 & u_nv_n & v_n^2\end{pmatrix}=(\mathbf{1}_n,\boldsymbol{X})$$

而令

$$\underline{\boldsymbol{B}}=\begin{pmatrix}\mu_1 & \mu_2 & \cdots & \mu_p\\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p}\\ \cdots & \cdots & \cdots & \cdots\\ \beta_{51} & \beta_{52} & \cdots & \beta_{5p}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{B}\end{pmatrix}.$$

在此例中,数据阵 $\underline{\boldsymbol{X}}$ 中的每一列不是代表一个独立变量,而是分别代表两个变量 u,v 的零次项,一次项,平方项和乘积项. 可见,即使当自变量与因变量之间的关系不是线性函数时,模型 18.4.2 仍可以用.

在模型 18.4.2 中,当 $p=1$ 时就是一元统计中的多重回归问题(在这里我们不用多元回归这个名词,以同 $p>1$ 时的多元回归相区别). 这时,熟知的最小二乘法是通过最小化残差平方和 $(\boldsymbol{Y}-\underline{\boldsymbol{X}}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y}-\underline{\boldsymbol{X}}\boldsymbol{\beta})$ 来求参数向量的估计. 在 $p>1$ 时,最小二乘法可定义为求解下述的极小值问题

$$\min_{\underline{\boldsymbol{B}}}\operatorname{tr}\{(\boldsymbol{Y}-\underline{\boldsymbol{X}}\,\underline{\boldsymbol{B}})^{\mathrm{T}}(\boldsymbol{Y}-\underline{\boldsymbol{X}}\,\underline{\boldsymbol{B}})\}. \tag{18.29}$$

同一元统计的结论相似,式(18.29)的解 $\hat{\underline{\boldsymbol{B}}}$ 满足下述的正规方程(normal equation)

$$\underline{X}^{\mathrm{T}}\underline{X}\hat{\underline{B}} = \underline{X}^{\mathrm{T}}Y. \tag{18.30}$$

在回归问题中，当 n 足够大时($n>q+1$)，我们总假定

$$\mathrm{rk}(\overline{X}) = q + 1. \tag{18.31}$$

这时正规方程有唯一解

$$\hat{\underline{B}} = (\underline{X}^{\mathrm{T}}\underline{X})^{-1}\underline{X}^{\mathrm{T}}Y. \tag{18.32}$$

为进一步将(18.32)式展开，我们给出下述的表达式

$$S_{xx} = X^{\mathrm{T}}\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)X, \tag{18.33}$$

$$S_{yy} = Y^{\mathrm{T}}\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)Y,$$

$$S_{xy} = X^{\mathrm{T}}\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)Y = S_{yx}^{\mathrm{T}},$$

$$\bar{y} = \frac{1}{n}\mathbf{1}^{\mathrm{T}}Y,\ \bar{x} = \frac{1}{n}\mathbf{1}^{\mathrm{T}}X.$$

其中 I 为 n 阶单位阵，$\mathbf{1}$为$\mathbf{1}_n$ 的简略表达式. 在(18.33)的记号下，我们有

$$\hat{\mu} = \bar{y} - \bar{x}\hat{B}, \tag{18.34}$$

$$\hat{B} = S_{xx}^{-1}S_{xy}.$$

利用预报方程(18.28)及上式，可得到 Y 的预报值为

$$\hat{Y} = (\mathbf{1}X)\begin{pmatrix}\hat{\mu}\\ \hat{B}\end{pmatrix} = \mathbf{1}\bar{y} + \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)X\hat{B}. \tag{18.35}$$

定义18.4.4 $\hat{\varepsilon} = Y - \hat{Y}$ 称为**残差**(residuals)，$Q = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} = (Y - \hat{Y})^{\mathrm{T}}(Y - \hat{Y})$称为残差阵.

容易推出，残差阵 Q 有表达式

$$Q = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}. \tag{18.36}$$

利用最小二乘估计$\hat{\underline{B}}$ 和残差阵 Q 可以给出参数阵$\underline{B}$ 和方差阵 Σ 的无偏估计.

定理18.4.5 在模型18.4.2下

(1) 正规方程(18.30)的解$\hat{\underline{B}}$ 为$\underline{B}$ 的无偏估计，即

$$E(\underline{\hat{B}}) = \underline{B}.$$

(2) $Q/(n-q-1)$为方差阵Σ的无偏估计，即

$$E(Q/(n-q-1)) = \Sigma.$$

定理18.4.6 记 $S_{xx}^{-1}=(S^{ij})$，$\Sigma=(\sigma_{ij})$，又记

$$B = (b_1, \cdots, b_p) = \begin{pmatrix} b_{(1)} \\ \vdots \\ b_{(q)} \end{pmatrix}，则$$

(1) $V(\hat{\mu}) = \left(\frac{1}{n} + \bar{x} S_{xx}^{-1} \bar{x}^{T}\right) \Sigma.$

(2) $\mathrm{cov}(\hat{b}_{(i)}^{T}, \hat{b}_{(j)}^{T}) = S^{ij}\Sigma, \quad i,j=1,2\cdots,q.$

(3) $\mathrm{cov}(\hat{b}_i, \hat{b}_j) = \sigma_{ij} S_{xx}^{-1}, \quad i,j=1,2,\cdots,p.$

为简化计算过程和便于分析，我们可事先使数据阵X中心化，即假定下述条件满足：

$$\bar{x} = \frac{1}{n} \mathbf{1}^{T} X = 0. \tag{18.37}$$

这时我们有下述结果

推论18.4.7 在条件18.37下有

(1) $S_{xx}=X^{T}X$，$S_{xy}=X^{T}Y=S_{yx}^{T}$.

(2) $\hat{\mu}=\bar{y}$ 且 $\hat{\mu}$ 与 $\hat{B}$ 不相关.

(3) $Q=Y^{T}Y-n\hat{\mu}^{T}\hat{\mu}-\hat{B}^{T}\hat{B}$.

在模型18.4.2的基础上，可进一步假设误差项服从正态分布，便得到

模型18.4.8 正态模型

$$\begin{cases} Y = (1, X)\begin{pmatrix} \mu \\ B \end{pmatrix} + E = \underline{X}\ \underline{B} + E, \\ E \text{ 的各行独立同 } N_p(0, \Sigma) \text{ 分布.} \end{cases}$$

定理18.4.9 在模型18.4.2及 $\mathrm{rk}(\underline{X})=q+1$的条件下有

(1) $\underline{B}$ 的极大似然估计为最小二乘估计$\underline{\hat{B}}$，Σ的极大似然估计

为$\hat{\boldsymbol{\Sigma}}=\frac{1}{n}\boldsymbol{Q}$.

(2) $\hat{\boldsymbol{B}}$ 服从正态分布且与 $\boldsymbol{Q}$ 独立.

(3) $\boldsymbol{Q}\sim W_p(n-q-1,\boldsymbol{\Sigma})$.

18.4.2 相关

回归这个概念可以说是从条件期望的角度来反映一个(或一些)随机变量对另一个(或一些)随机变量的依赖性(dependence). 另一方面可以从各种类型的(条件)方差和(条件)标准差之比的角度来看随机变量之间的依赖性和相依性(interdependence),这就是熟知的相关系数概念的推广. 本小节介绍一些在多元分析中有用的相关概念,以便使读者能更好地理解回归分析模型.

在理论回归模型18.4.1中,设 p 等于1,假定 x 是随机变量,并记 $\boldsymbol{B}=\boldsymbol{b}$:$q$ 维列向量,则我们有下述结论

定理18.4.10

(1) $V(y)=V(x\boldsymbol{b})+V(\boldsymbol{\varepsilon})=V(x\boldsymbol{b})+\boldsymbol{\Sigma}$

(2) $$\operatorname{cor}^2(y,x\boldsymbol{b})=\max_{\beta}\operatorname{cor}^2(y,x\boldsymbol{\beta})=\frac{V(x\boldsymbol{b})}{V(y)}.$$

由此定理可看出:第一,y 的方差由 $x\boldsymbol{b}$ 之方差与 $\boldsymbol{\varepsilon}$ 的方差之和构成,而 y 与 $x\boldsymbol{b}$ 之相关系数的平方即为 $V(x\boldsymbol{b})$与 $V(y)$之比. 第二,在所有 x 的线性组合中,$x\boldsymbol{b}$ 与 y 的相关系数(绝对值)最大. 因此,可以用 $V(x\boldsymbol{b})/V(y)$作为 x 解释 y 的"能力"大小的一个度量. 此比值越大,则说明 x 解释 y 的能力越大,否则说明 x 解释 y 的能力越小. 在实际中,我们用 $V(x\boldsymbol{b})/V(y)$的样本形式. 在回归的具体模型18.4.2中,设 $p=1$,记 $\boldsymbol{B}=\boldsymbol{b}$,用 $\boldsymbol{b}$ 的最小二乘估计 $\hat{\boldsymbol{b}}$ 代 $\boldsymbol{b}$,则 $X\hat{\boldsymbol{b}}$ 的样本方差为 $\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}$,记 $\boldsymbol{Y}$ 的样本方差为 S_{yy}.

定义18.4.11 y 对于 $\boldsymbol{x}$ 的**复相关系数**(multiple correlation coefficient)为

$$R = \sqrt{\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}/S_{yy}}.$$

当 R 较大时，回归模型理想（$\boldsymbol{x}$ 能较好地解释 y），否则不理想.

定理18.4.12

（1）$R^2=1-(Q/S_{yy})$，其中 Q 为残差.

（2）在正态模型18.4.8之下，

$$\frac{q}{n-q-1}\frac{1-R^2}{R^2}\sim F(n-q-1,q).$$

当 $p=2$时，在理论模型18.4.1中记 $\boldsymbol{y}=(y_1,y_2)$，$\boldsymbol{B}=(\boldsymbol{b}_1,\boldsymbol{b}_2)$. 由于模型的影响，$y_1$与 y_2之间的相关性有 $\boldsymbol{x}$ 变量的作用在里面. 我们要进一步考察在扣除了 $\boldsymbol{x}$ 的影响之后，y_1与 y_2之间的相关性，实际上是关心

$$\mathrm{cor}(y_1 - xb_1, y_2 - xb_2)$$

的值.

在样本模型18.4.2中设条件(18.37)成立，记 $\boldsymbol{Y}=(\boldsymbol{y}_1,\boldsymbol{y}_2)$，$\hat{\boldsymbol{B}}=(\hat{\boldsymbol{b}}_1,\hat{\boldsymbol{b}}_2)$为 $\boldsymbol{B}$ 的最小二乘估计，又记

$$\hat{\boldsymbol{e}}_i = \boldsymbol{y}_i - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{y}_i - \boldsymbol{X}\hat{\boldsymbol{\beta}}_i \quad i=1,2. \qquad (18.38)$$

定义18.4.13 变量 y_1，y_2关于 $\boldsymbol{x}$ 的**偏相关系数**(partial correlation coefficient)定义为

$$R_{12\cdot x} = \hat{\boldsymbol{e}}_1^{\mathrm{T}}\hat{\boldsymbol{e}}_2/(\|\hat{\boldsymbol{e}}_1\|\cdot\|\hat{\boldsymbol{e}}_2\|).$$

在实际问题中，经常会碰到这样的情况：两个随机变量之间相关系数很大，但在扣除了其它一些变量的影响后，偏相关系数却很小，这说明两个变量之间的相关性是由其它一些变量所产生的. 相反的情形也会发生.

定理18.4.14 设 $\boldsymbol{Q}(2\times 2)$为残差阵，记 $\boldsymbol{R}$ 为相应的相关阵，即 $\boldsymbol{R}$ 的元素 $r_{ij}=q_{ij}/(q_{ii}q_{jj})^{\frac{1}{2}}$，则 $R_{12\cdot x}$即为 R 的非对角元素.

回顾4.2.1中复相关系数的定义，是考虑 $p=1$时，即1个变量

对多个变量的相依性，$R^2=1-\{(y-\hat{y})^{\mathrm{T}}(y-\hat{y})/(y-\bar{y})^{\mathrm{T}}(y-\bar{y})\}$. 现在我们要将此概念推广到 $p>1$的情形，即同时考虑多个变量对多个变量的相依性问题. 自然由上述 R^2的表达式，我们推广到下述的矩阵.

$$R^2=\boldsymbol{I}-\{(\boldsymbol{Y}-\mathbf{1}\bar{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{Y}-\mathbf{1}\bar{\boldsymbol{y}})\}^{-1}(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^{\mathrm{T}}(\boldsymbol{Y}-\hat{\boldsymbol{Y}}), \quad (18.39)$$

其中 $\bar{\boldsymbol{y}}$ 为 $\boldsymbol{Y}$ 的均值行向量，$\hat{\boldsymbol{Y}}$ 为 $\boldsymbol{Y}$ 的预报值.

由 R^2出发可以有几种不同的方式定义广义相关系数. 我们给出下面两个不等价的定义.

定义18.4.15 下述的两个量都定义为 p 维变量 $\boldsymbol{y}$ 关于 q 维向量 x 的广义相关系数：

(1) $R_T=\left\{\dfrac{1}{p}\mathrm{tr}(\boldsymbol{R})^2\right\}^{1/2}$.

(2) $R_D=\{|\boldsymbol{R}^2|\}^{1/2}$.

定理18.4.16

(1) $\boldsymbol{R}^2=\boldsymbol{I}-\boldsymbol{S}_{yy}^{-1}(\boldsymbol{S}_{yy}-\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy})=\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}$.

(2) 设 $\lambda_1,\cdots,\lambda_p$ 为 $\boldsymbol{R}^2$的特征根，则$1\geqslant\lambda_i\geqslant0, i=1,\cdots,p$，且

$$R_T^2=\frac{1}{p}\sum_1^p\lambda_i, R_D^2=\prod_1^p\lambda_i.$$

由上述定义可见 R_T^2 与 R_D^2 有一些相同的性质，例如，都在0与1之间，且当接近1时，表示 $\boldsymbol{y}$ 对 $\boldsymbol{x}$ 有较大的依赖性；反之当接近零时，表示 $\boldsymbol{y}$ 基本上不受 $\boldsymbol{x}$ 的影响. 但有一个重要的不同点. 即 $R_T^2\geqslant R_D^2$，=号成立当且仅当 $\lambda_1=\lambda_2=\cdots=\lambda_p$. 此外，当 $\lambda_1,\cdots,\lambda_p$ 中有一个为零时 $R_D^2=0$；而 $R_T^2=0$，当且仅当 $\lambda_1=\cdots=\lambda_p=0$. 由此看来，用 R_T 作广义相关系数比较合理.

18.4.3 假设检验

在多元回归分析的假设检验问题中，我们总是从正态回归模型出发，要检验的假设可归结为对参数阵 B 的一个线性假定

$$H_0: \boldsymbol{RB} = 0, \tag{18.40}$$

其中 $\boldsymbol{R}$ 为 $s\times q$ 的常数阵,$\mathrm{rk}(\boldsymbol{R})=s$. 所采用的检验方法是似然比检验. 我们有下面的一个一般的结论.

定理18.4.17 在假设18.40之下,用 $\boldsymbol{Q}_0$记当 H_0成立时得到的残差阵,用 $\boldsymbol{Q}_1$记当 H_0不成立时得到的残差阵,则有

(1) 似然比统计量有形式

$$\lambda = [|\boldsymbol{Q}_1|/|\boldsymbol{Q}_0|]^{\frac{n}{2}}.$$

(2) $U=\lambda^{\frac{2}{n}}=|\boldsymbol{Q}_1|/|\boldsymbol{Q}_0|\sim\Lambda(p,n-q-1,s)$.

因此对给定的显著性水平 α,似然比检验的形式为

$$\text{当 } U < \Lambda_\alpha(p,n-q-1,s) \text{ 时否定 } H_0; \tag{18.41}$$

否则就肯定 H_0.

其中 $\Lambda_\alpha(p,n-q-1,s)$为 $\Lambda(p,n-q-1,s)$的下 α 分位点. 即

$$P\{\Lambda(p,n-q-1,s) < \Lambda_\alpha(p,n-q-1,s)\} = \alpha. \tag{18.42}$$

下面我们给出几个有用的特例. 首先考虑检验自变量 $\boldsymbol{x}$ 的全体是否对因变量 $\boldsymbol{y}$ 有影响的问题,也就是要检验假设

$$H_0: B = 0. \tag{18.43}$$

这相当于在18.40中取 $\boldsymbol{R}=\boldsymbol{I}_q$.

定理18.4.18 对假设(18.43)有

(1) $\boldsymbol{Q}_1=\boldsymbol{S}_{yy}-\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}$,$\boldsymbol{Q}_0=\boldsymbol{S}_{yy}$.

(2) $U=|\boldsymbol{Q}_1|/|\boldsymbol{Q}_0|=|\boldsymbol{S}_{yy}-\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}|/|\boldsymbol{S}_{yy}|$

$\quad=|\boldsymbol{I}-\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}|\sim\Lambda(p,n-q-1,q)$.

因此当 $U<\Lambda_\alpha(p,n-q-1,q)$时否定 H_0;否则肯定 H_0.

下面我们给出一个记号

$$\boldsymbol{S}_{yy\cdot x} = \boldsymbol{S}_{yy} - \boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}. \tag{18.44}$$

用这个记号可简化 $\boldsymbol{Q}_1$的表达式. 另外这个记号还可以有一些变形,在下面是很有用的.

当因变量的个数 $p=1$时,定理18.4.18中给出的基于 Λ 分布

的检验可转化为 F 检验.

定理18.4.19 当 $p=1$时，在假设(18.41)下

(1) $U=1-R^2$，其中 $R^2=\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}/\boldsymbol{S}_{yy}$为复相关系数.

(2) $T^2=\frac{1-U}{U}=\frac{R^2}{1-R^2}\sim\frac{q}{n-q-1}F(q,n-q-1)$.

由于T^2是U的单调下降函数，因此似然比检验有下列形式

当$\frac{n-q-1}{q}T^2>F_\alpha(q,n-q-1)$则否定$H_0$； (18.45)

否则便肯定H_0.

其中$F_\alpha(q,n-q-1)$为$F(q,n-q-1)$的上α分位数. 即

$$P\{F(q,n-q-1)>F_\alpha(q,n-q-1)\}=\alpha. \quad (18.46)$$

注 请读者注意，当用Λ分布检验时，临界值总是用下α分位数，而用F分布检验时，临界值总是用上α分位数.

其次，当假设18.41被拒绝时，我们要进一步检验自变量中的一部分是否对因变量起作用. 为表达方便起见，不妨设为检验q个自变量中的后s个. 记

$$\underset{n\times q}{\boldsymbol{X}}=(\underset{q-s}{\boldsymbol{X}_1},\underset{s}{\boldsymbol{X}_2}),\ \underset{q\times p}{\boldsymbol{B}}=\begin{pmatrix}\boldsymbol{B}_{(1)}\\ \boldsymbol{B}_{(2)}\end{pmatrix}\begin{matrix}q-s\\ s\end{matrix}. \quad (18.47)$$

要检验的假设是

$$H_0:\ \boldsymbol{B}_{(2)}=0. \quad (18.48)$$

定理18.4.20 在假设(18.42)之下有

(1) $Q_1=\boldsymbol{S}_{yy\cdot x},Q_0=\boldsymbol{S}_{yy\cdot x_1}$.

其中$S_{yy\cdot x_1}$是在$S_{yy\cdot x}$的表达式中(见(18.44))用$\boldsymbol{X}_1$代替$\boldsymbol{X}$得到的.

(2) $U=|\boldsymbol{Q}_1|/|\boldsymbol{Q}_0|\sim\Lambda(p,n-q-1,s)$.

假设(18.48)的一个特例是当$s=1$，即检验最后一个自变量是否对模型有影响的问题. 我们在(18.47)中记$u=\boldsymbol{X}_2,\boldsymbol{b}_{(u)}=\boldsymbol{B}_{(2)}$.

定理18.4.21 当$S=1$时，在假设(18.46)之下有

(1) $U=|\boldsymbol{Q}_1|/|\boldsymbol{Q}_0|=\frac{1}{1+d\hat{\boldsymbol{B}}_{(2)}\boldsymbol{Q}_1^{-1}\hat{\boldsymbol{B}}_{(2)}^{\mathrm{T}}}$.

其中 $d=S_{uu\cdot x_1}$，$\hat{B}_{(2)}=d^{-1}(S_{uy\cdot x_1})$为在原模型下 $B_{(2)}$的估计. $Q_1=S_{yy\cdot x}$.

(2) $T^2=\frac{1-U}{U}=d\hat{B}_{(2)}Q_1^{-1}\hat{B}_{(2)}^{\mathrm{T}}\sim\frac{p}{n-q-p}F(p,n-q-p)$.

注 $S_{uy\cdot x_1}=S_{uy}-S_{ux_1}S_{x_1x_1}^{-1}S_{x_1y}$.

这式是(18.44)的进一步一般化.

18.4.4 剔除变量与逐步回归

在多元回归模型中，通常在建模时应采集尽可能多的数据，尽量把可能有关的自变量和因变量都包括进来. 但在进行分析时便会发现并不是所有的自变量对所有的因变量都有作用. 通常可能的情形是：因变量分成若干组，每组因变量对应着一组自变量构成一个回归模型；而不同组的因变量之间没有什么太大的关系，这与一元的多重回归模型不同. 在那里因变量只有1个，不存在分组的问题. 逐步回归的任务只是要找出一组对因变量有显著影响的自变量而把多余的自变量剔除. 而在多元的逐步回归中面临着双重筛选的任务. 多元逐步回归的理论依据便是上一小节中介绍的假设检验原理. 下面我们不加证明地给出一个双重筛选的逐步回归算法.

首先作预备工作，设因变量共有 p 个，自变量共有 q 个，作了 n 次独立观测. 记因变量数据阵为 $Y(n\times p)$，自变量数据阵为 $X(n\times q)$

$$Y=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix},\quad X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix}.$$

其中 Y 的第 k 行表示对 p 个因变量的第 k 次独立观测；X 的第 k 行表示对 q 个自变量的第 k 次观测，$k=1,\cdots,n$. 先按照下述公式

计算样本均值.

$$\bar{y}_i = \frac{1}{n}\sum_{k=1}^{n} y_{ki}, \quad i = 1,\cdots,p,$$

$$\bar{x}_j = \frac{1}{n}\sum_{k=1}^{n} x_{kj}, \quad j = 1,\cdots,q.$$

然后按下述公式计算样本方差阵.

$$S_{yy}(i,j) = \sum_{k=1}^{n}(y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j) \quad i,j = 1,\cdots,p,$$

$$S_{yx}(i,j) = \sum_{k=1}^{n}(y_{ki} - \bar{y}_i)(x_{kj} - \bar{x}_j) \quad i = 1,\cdots,p, j = 1,\cdots,q.$$

$$S_{xx}(i,j) = \sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad i,j = 1,\cdots,q.$$

$$S_{yy}(p \times p) = [S_{yy}(i,j)].$$

$$S_{yx}(p \times q) = [S_{yx}(i,j)] = S_{xy}^{T}.$$

$$S_{xx}(q \times q) = [S_{xx}(i,j)].$$

在逐步回归中，一个基本的计算过程是对样本方差矩阵反复进行一种矩阵变换，其定义如下.

定义18.4.22 设 $\boldsymbol{A} = (a_{kl})$ 为 $n \times m$ 阵，对 $a_{ij} \neq 0$，$\boldsymbol{A}$ 的一个变换 $\boldsymbol{A}^* = T_{ij}(\boldsymbol{A})$ 也是一个 $n \times m$ 矩阵，其元素 a_{kl}^* 定义为

$$a_{kl}^* = \begin{cases} \dfrac{1}{a_{ij}}, & \text{当 } k = i, l = j, \\ a_{il}/a_{ij}, & \text{当 } k = i, l \neq j, \\ -a_{kj}/a_{ij}, & \text{当 } k \neq i, l = j, \\ a_{kl} - (a_{il}a_{kj})/a_{ij}, & \text{当 } k \neq i, l \neq j. \end{cases}$$

变换 T_{ij} 称为矩阵的 (i,j) 消去变换.

性质18.4.23 (i,j) 消去变换的性质.

(1) $T_{ij}(T_{ij}(\boldsymbol{A})) = \boldsymbol{A}$.

(2) 若 $i \neq k$，且 $j \neq l$，则 $T_{ij}(T_{kl}(\boldsymbol{A})) = T_{kl}(T_{ij}(\boldsymbol{A}))$.

(3) 若 $\boldsymbol{A}$ 为 $n\times n$ 方阵，记 $\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}\begin{matrix}(r)\\(n-r)\end{matrix}$，列分块为 (r)，$(n-r)$.

设 $\boldsymbol{A}_{11}^{-1}$ 存在，则

$$T_{11}(T_{22}(\cdots T_{rr}(\boldsymbol{A})\cdots)) = \begin{pmatrix}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{pmatrix}.$$

记

$$\boldsymbol{S} = \begin{pmatrix}\boldsymbol{S}_{xx} & \boldsymbol{S}_{xy}\\ \boldsymbol{S}_{yx} & \boldsymbol{S}_{yy}\end{pmatrix}\begin{matrix}(q)\\(p)\end{matrix}. \tag{18.49}$$

（列分块为 (q)，(p)）

对 $\boldsymbol{S}$ 作 $(1,1),(2,2),\cdots,(q,q)$ 消去变换，由18.4.23易见

$$\boldsymbol{S}\rightarrow\begin{pmatrix}\boldsymbol{S}_{xx}^{-1} & \boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}\\ -\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1} & \boldsymbol{S}_{yy}-\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}\end{pmatrix}. \tag{18.50}$$

据(18.34)和(18.36)可知，变换后的矩阵右上角 $\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}=\hat{\boldsymbol{B}}$，为回归系数阵 $\boldsymbol{B}$ 的估计；右下角为 $\boldsymbol{S}_{yy\cdot x}=\boldsymbol{Q}$，即残差阵.

在双重筛选的逐步回归中，使用两个工作矩阵 $\boldsymbol{S}_1=\boldsymbol{S}_2=\boldsymbol{S}$（见(18.49)）. 记

$$\boldsymbol{S}_1=\begin{pmatrix}\boldsymbol{S}_{11}\\ \cdots\cdots\\ \boldsymbol{S}_{12}\end{pmatrix},$$

$$\boldsymbol{S}_{11}=(\boldsymbol{S}_{xx}\ \vdots\ \boldsymbol{S}_{xy}),\boldsymbol{S}_{12}=(\boldsymbol{S}_{yx}\ \vdots\ \boldsymbol{S}_{yy}).$$

基本的计算过程如下：当第 i 个自变量入选，则对 $\boldsymbol{S}_{11}$ 和 $\boldsymbol{S}_2$ 同时作 (i,i) 消去变换，而 $\boldsymbol{S}_{12}$ 不变；当第 j 个因变量入选则对 $\boldsymbol{S}_{12}$ 和 $\boldsymbol{S}_2$ 同时作 (j,j) 消去变换，而 $\boldsymbol{S}_{11}$ 不变.

假定已运算了若干步，入选了 l 个因变量，k 个自变量，相应的工作矩阵已变到 $\boldsymbol{S}_1^*=\begin{pmatrix}\boldsymbol{S}_{11}^*\\ \cdots\cdots\\ \boldsymbol{S}_{12}^*\end{pmatrix}$ 及 $\boldsymbol{S}_2^*$. 记 $\boldsymbol{S}_{11}^*,\boldsymbol{S}_{12}^*,\boldsymbol{S}_2^*$ 中的元素为 $\boldsymbol{S}_{11}^*(s,t),\boldsymbol{S}_{12}^*(s,t)$ 及 $\boldsymbol{S}_2^*(s,t)$，根据下述准则对变量进行入选或

剔除.

算法18.4.24

（1）对未入选的自变量 x_i 计算

$$T_i^2=(S_{11}^*(i,i)/S_2^*(i,i))-1.$$

在所有未入选的自变量所对应的 T_i^2 中找一个最大的，设为 $T_{i_1}^2$. 对给定的阈值 F_{in}^x，检验下列不等式

$$\frac{n-k-l-1}{l}T_{i_1}^2>F_{in}^x.$$

若不等式成立，则入选自变量 x_{i_1}，即作变换.

$$T_{i_1i_1}(S_{11}^*),T_{i_1i_1}(S_2^*).$$

然后转入(2)，否则不入选任何自变量，即 S_{11}^*,S_{12}^*,S_2^* 保持不变，转入(2).

（2）对已入选的自变量 x_i 计算

$$T_i^2=(S_2^*(i,i)/S_{11}^*(i,i))-1.$$

在所有已入选的自变量所对应的 T_i^2 中找一个最小的，设为 $T_{i_2}^2$，对给定的阈值 F_{out}^x 检验下列不等式

$$\frac{n-k-l}{l}T_{i_2}^2<F_{out}^x.$$

若不等式成立，则剔除自变量 x_{i_2}，即作变换

$$T_{i_2i_2}(S_{11}^*),\ T_{i_2i_2}(S_2^*).$$

然后回到(1). 若不等式不成立，则不剔除任何自变量，即保持 S_{11}^*,S_{12}^*,S_2^* 不变，回到(1).

当(1)和(2)中既没有自变量入选，又没有自变量剔除时就转入(3).

（3）对未入选的因变量 y_j，计算

$$T_j^2=[S_{12}^*(j,j)/S_2^*(j,j)]-1.$$

在所有未入选的因变量所对应的 T_j^2 中选一个最大的. 设为

$T^2_{j_1}$. 对给定的阈值 F^y_{in}，检验下列不等式

$$\frac{n-k-l-1}{k}T^2_{j_1} > F^y_{in}.$$

若不等式成立，则入选因变量 y_{j_1}，即作变换

$$T_{q+j_1,q+j_1}(\boldsymbol{S}^*_{12}),\ T_{q+j_1,q+j_1}(\boldsymbol{S}^*_2).$$

然后回到(1)，否则不入选任何因变量，即保持 $\boldsymbol{S}^*_{11},\boldsymbol{S}^*_{12},\boldsymbol{S}^*_2$ 不变，转入(4).

(4) 对已入选的因变量 y_j，计算

$$T^2_j = [S^*_2(j,j)/S^*_{12}(j,j)] - 1.$$

在所有已入选的因变量所对应的 T^2_j 中找一个最小的. 设为 $T^2_{j_2}$. 对给定的阈值 F^y_{out}，检验下列不等式

$$\frac{n-k-l}{k}T^2_{j_2} < F^y_{out}.$$

若不等式成立，则剔除因变量 y_{j_2}，即作变换

$$T_{q+j_2,q+j_2}(S^*_{12}),\ T_{q+j_2,q+j_2}(S^*_2).$$

然后回到(1). 否则，不剔除任何自变量，即 $\boldsymbol{S}^*_{11},\boldsymbol{S}^*_{12}$与 $\boldsymbol{S}^*_2$ 保持不变. 逐步回归停止. 此时，自变量和因变量在上述筛选的过程中都不再发生变化.

这时，在 $\boldsymbol{S}^*_{11}$的右边对应已入选的 k 个自变量的行和已入选的 l 个因变量的列上的元素就是这 k 个自变量对 l 个因变量的回归系数(构成一个 $k\times l$ 矩阵). 常量项 $\boldsymbol{\mu}$ 的估计可用公式(18.32)得到，但须注意，这时的 $\bar{\boldsymbol{y}},\bar{\boldsymbol{x}}$ 及 $\hat{\boldsymbol{B}}$ 只与入选的自变量与因变量有关. 如要得到残差阵，则还需对 $\boldsymbol{S}^*_2$ 关于已入选的因变量作(j,j)消去变换，则得到的相应的 $l\times l$ 矩阵(位于 $\boldsymbol{S}^*_2$ 的右下角)即为残差阵.

在 $\boldsymbol{S}$ 阵中将已入选的因变量所对应的行和列划掉，得到的是未入选的因变量对全体自变量的 $\boldsymbol{S}$ 阵. 由此出发重新进行逐步回归就得到一组新的回归方程.

这里因变量是与(1)中选中的因变量完全不同，但自变量可以

是重选的. 如此往复,直至全体因变量分成 S 组,每一组对应着一组自变量或者没有自变量相对应. 最后一种情形说明那些因变量不能用任何自变量来解释.

最后讲几个有关的技术细节.

首先是阈值 F_{in}^{x}, F_{out}^{x}, F_{in}^{y}和 F_{out}^{y}的选取问题,从理论上讲,这些阈值就是 F 分布的分位数. 严格来讲,应该按照给定的检验水平和相应 F 分布的自由度来给出. 但这样做势必使计算过程复杂化. 一个简单的做法是令 $F_{in}^{x}=F_{out}^{x}\sim3-10$, $F_{in}^{y}=F_{out}^{y}\sim1-3$.

其次是逐步回归的初始状态的问题,即事先指定哪些自变量和因变量已入选. 有几种考虑可供选择. 第一,入选方差最大的因变量和与之相关系数最大的自变量;第二,根据实际背景入选那些用户认为是最重要的变量;第三,入选全体因变量. 不管基于哪种考虑,假定出发点是已入选的自变量 $x_{i_1}, x_{i_2}, \cdots, x_{i_k}$,及因变量 y_{j_1}, $y_{j_2}, \cdots, y_{j_l}$,则对 $\boldsymbol{S}_{11}$作$(i_1, i_1), \cdots, (i_k, i_k)$消去变换,对 $\boldsymbol{S}_{12}$作(j_1, j_1), $\cdots, (j_l, j_l)$消去变换,对 $\boldsymbol{S}_2$同时作上述二组消去变换,然后以此为出发点进行逐步回归. 需要说明的是,由不同的初始状态进行逐步回归,其结果可能是不一样的. 因此,如果有足够的经费和时间的话,可以从几个不同的初始状态出发,将得到的结果进行比较,从中选一个较为理想的结果.

最后,在对 $\boldsymbol{S}_{11}$, $\boldsymbol{S}_{12}$和 $\boldsymbol{S}_2$进行消去变换时,如果出现了对角线为零的元素,则说明产生了严格线性相关的情形,此时可将相应的变量删掉,再继续进行.

18.5 判别分析

18.5.1 序言

判别分析(discriminant analysis)是多元分析中又一类应用很

广的方法，本节介绍一些最常用的判别方法. 首先，我们给出基本概念.

假如一个医生对前来求治的病人进行诊断、化验，得到一系列生理指标，如体温，血压，X 光透视结果等等. 医生要根据得到的这一组数据来判断该病人是生了哪一种病：感冒，肺炎或其它的病？这是一个判别问题. 又例如在地质勘探中，需要从岩石标本的各种特征来判别该岩石样本是取自哪一种地质构造. 考古学家要根据出土文物的各种特征来判别这些出土文物是来自哪一历史时代或哪一种文化形态等等. 以上所举的例子都可以用统计学的判别分析的方法来处理. 总之，判别分析在生产和科研的很多领域里都可以找到自己的用途.

用数学语言可以将判别问题表述如下：设有一个总体 $\mathscr{P}$，其中包含了 m 个子体 $\Pi_1, \Pi_2, \cdots, \Pi_m$，每个子体在总体中所占的比例为 $\pi_1, \pi_2, \cdots, \pi_m$，$\sum\limits_{1}^{m}\pi_i = 1$. 又设每个子体本身有概率分布（离散变量场合）或密度函数（连续变量场合）$f_i(\boldsymbol{x})$，$i=1,\cdots,m$，其中 $\boldsymbol{x}$ 为 p 维变量. 对从总体 $\mathscr{P}$ 中取出的一个样本 $\boldsymbol{x}$，我们要判断它来自哪个子体. 具体作法如下.

定义18.5.1 将 p 维欧氏空间 R^p 划分为 m 个互不相交的区域 $R_1, R_2, \cdots, R_m$，即 $R_i \cap R_j = \varnothing$，当 $i \neq j$；$\bigcup\limits_{1}^{m} R_i = R^p$. 对给定的 $\boldsymbol{x} \in R^p$，当 $\boldsymbol{x} \in R_i$ 时，就判定 $\boldsymbol{x}$ 是取自子体 Π_i，$i=1,\cdots,m$. 这个 R^p 的**划分**(partition)便称为是一个**判别法则**(discriminant rule).

显然，要解决一个判别问题就要解决如何确定一个“好的”判别法则的问题. 这里“好的”概念可以是基于某种理论准则，也可以是基于某种直观可行的想法. 但是，无论何种判别法则都要面临发生错判的问题. 假定样本 x 是取自子体 Π_i，而根据判别法则 $\mathscr{D}$ 判定 x 属于 Π_j，于是就发生了错误. 错判概率为

$$P_{\text{err}}(j|i) = \int_{R_j} f_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad i \neq j. \tag{18.51}$$

这里积分号表示重积分(对连续变量)或求和(对离散变量).于是,对一个判别法则 $\mathscr{D}$ 来说,其总的平均错判概率为

$$\begin{aligned} P_{\text{err}} &= \sum_{i=1}^{m} P(\boldsymbol{x} \text{ 取自 } \Pi_i) \cdot P(\boldsymbol{x} \overline{\in} R_i | x \text{ 取自 } \Pi_i) \\ &= \sum_{i=1}^{m} \pi_i \sum_{j\neq i} P_{\text{err}}(j|i) = \sum_{i=1}^{m} \Pi_i \sum_{j\neq i} \int_{R_j} f_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{18.52}$$

不论用何种判别法则,我们都希望错判概率尽可能小;在实际问题中,由于 $f_i(x)$ 都是未知或部分未知的,还存在对错误概率进行估计的问题.

18.5.2 分布已知时的判别法则

假定 m 个子体 $\Pi_1, \cdots, \Pi_m$ 在总体 $\mathscr{P}$ 中所占的比例 $\pi_1, \cdots, \pi_m$ 已知,又假定各子体本身的分布 $f_1(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x})$ 也已知,则最常用的判别方法是贝叶斯(Bayes)判别,其定义如下

定义 18.5.2 一个判别法则,即 R^p 的一个划分 $R_1, \cdots, R_m$ 称为是贝叶斯判别法则,若它使(18.52)所定义的错判概率 P_{err} 达到最小.

简言之,贝叶斯判别法则就是**错判概率最小法则**.下面的定理给出了贝叶斯法则的具体形式.

定理 18.5.3 贝叶斯判别法则是:当 $\Pi_i f_i(\boldsymbol{x}) \geqslant \Pi_j f_j(\boldsymbol{x}), j = 1, \cdots, m$,则判定 $x \in \Pi_i$,即取

$$R_i = \{x: \pi_i f_i(\boldsymbol{x}) = \max_{1\leqslant j\leqslant m} \pi_j f_j(\boldsymbol{x})\} \quad i = 1, \cdots, m.$$

特别当 $m=2$ 时有

$$\begin{aligned} R_1 &= \left\{x: \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geqslant \frac{\pi_2}{\pi_1}\right\} \\ &= \left\{x: \log \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geqslant \log \frac{\pi_2}{\pi_1}\right\}, \end{aligned}$$

$$R_2 = R^p - R_1.$$

注：在上述结论中，对某个给定的样本 $\boldsymbol{x}$ 可能达到最大的 $\Pi_i f_i(\boldsymbol{x})$ 不止一个，这时可以事先规定将 $\boldsymbol{x}$ 划入某个 R_i，例如可以规定 $\boldsymbol{x}$ 划入标号较小的 R_i. 这种情形在离散场合常会遇到，而在连续场合则较少发生.

例 18.5.4 设 m 个子体为具有相同方差阵，不同均值的正态分布，即 $\Pi_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$，$i=1,\cdots,m$. 则 $\pi_i f_i(\boldsymbol{x}) = \pi_i C \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right\}$. 两边取对数，我们有

$$\log(\pi_i f_i) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) + \log\pi_i + C',$$

$$i = 1,\cdots,m. \tag{18.53}$$

忽略掉常数项和共同的 $-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ 项，要使 $\pi_i f_i(\boldsymbol{x})$ 达到最大，等价于使 $\left(\boldsymbol{x}-\frac{\boldsymbol{\mu}_i}{2}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \log\pi_i$ 达到最大. 因此贝叶斯法则为

$$R_i = \left\{\boldsymbol{x}: \left(\boldsymbol{x}-\frac{\boldsymbol{\mu}_i}{2}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \log\pi_i = \max_{1\leqslant j\leqslant m}\left[\left(\boldsymbol{x}-\frac{\boldsymbol{\mu}_j}{2}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j + \log\pi_j\right]\right\}, i=1,\cdots,m. \tag{18.54}$$

特别当 $m=2$，即只有两个子体时，容易推出贝叶斯判别法则为

$$R_1 = \left\{\boldsymbol{x}: \left(\boldsymbol{x}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\right)^{\mathrm{T}}\boldsymbol{a} > \log\frac{\pi_2}{\pi_1}\right\}, R_2 = R^p - R_1, \tag{18.55}$$

其中 $\boldsymbol{a} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

注 在(18.55)中，判别是依据 x 的一个函数 $h(\boldsymbol{x})$ 的大小来决定的，这函数便称为**判别函数**(discriminant function). 特别在此场合判别函数是线性的，便称为是线性判别函数，相应的判别法则便称为线性判别.

由贝叶斯判别法则可以引伸出一些别的判别法则，首先考虑

当$\{\pi_i\}$——子体占总体$\mathscr{P}$的比例(也可看成是属于某个Π_i的先验概率)不知道的场合,在这种场合,我们可以假定

$$\pi_1 = \pi_2 = \cdots = \pi_m. \tag{18.56}$$

这时贝叶斯判别法则变成了

$$R_i = \{x: f_i(x) = \max_{1\leqslant j\leqslant m} f_j(x)\}, \quad i = 1,\cdots,m. \tag{18.57}$$

这就是**最大似然判别法**(maximum likelihood discriminant).因此最大似然判别法是贝叶斯判别的特例.

我们进一步考虑发生错判所造成的损失,当$x\in\Pi_i$而判$x\in\Pi_j$所造成的损失与当$x\in\Pi_k$而判$x\in\Pi_l$所造成的损失可能是不同的.例如,当病人无病而医生诊断为有病,其结果不过是虚惊一场;而当病人有病医生却诊断为无病,结果可能是延误时机造成不可挽回的损失.因此,在某些问题中我们需要考虑错判损失问题.

定义 18.5.5 设$L(i|j)\geqslant 0, i,j=1,\cdots,m, i\neq j$为一组损失,即当$x\in\Pi_j$而判为$\in\Pi_i$时有损失$L(i|j)$,则总损失为

$$L = \sum_{i=1}^{m}\sum_{j\neq i} L(j|i)\pi_i \int_{R_j} f_i(x)\mathrm{d}x.$$

称使得总损失L达到最小的判别法则为损失最小判别法则.

注 在一些文献中定义损失最小判别法则为贝叶斯判别法则,这样我们在18.5.4中定义的贝叶斯判别法则就成了一个特例,即在$L(i|j)=$常数下的贝叶斯法则.这里的区别仅是名词使用上的不同而不存在实质上的区别.

最小损失判别法则有形式

$$R_i = \left\{x: \sum_{j\neq i}\pi_j f_j(x)L(i|j) = \min_{1\leqslant k\leqslant m}\sum_{j\neq k}\pi_j f_j(x)L(k|j)\right\},$$
$$i = 1,\cdots,m. \tag{18.58}$$

在实际问题中选择一组损失$L(i|j)$要根据具体问题的背景来给出,但在很多情形下,损失是难以数量化的,因此硬要给出一组损失就难免带上主观随意性.另一方面,对一个总平均错判概率

(P_{err})很小的判别法则来说，某个条件错误概率 $P_{err}(j|i)$（见(18.51)）却可能较大，特别是当 π_i 较小时更是如此，这种情形有时是要特别避免的. 这种情形常发生在诊断某些少见的疾病或探测某种稀有矿产中. 为此引进**最大错判概率最小判别**(minimax discriminant)的概念.

定义 18.5.6 设某一判别法则满足：最大错判概率 $\max\{P_{err}(i|j) \quad i\neq j\}$在所有判别法则中达到最小，则称此判别法则为最大错判概率最小法则，简称为 minimax 判别.

定理 18.5.7 minimax 判别法则有如下的形式：

$$R_i = \{\boldsymbol{x}: \alpha_i f_i(\boldsymbol{x}) > \max_{1\leqslant j\leqslant m} \alpha_j f_j(\boldsymbol{x})\}, \quad i = 1,\cdots,m,$$

其中 $\alpha_i \geqslant 0$, $\sum_1^m \alpha_i = 1$，且 $P_{err}(j|i) = P_{err}(k|l)$，对任何 $j\neq i$，及 $k\neq l$都成立.

18.5.3 分布含未知参数时的判别法则

在实际问题中，分布完全知道的情形几乎不存在. 本小节讨论分布部分未知的情形，即子体 Π_i 的分布 $f_i(\boldsymbol{x},\boldsymbol{\theta}_i)$中有未知参数 $\boldsymbol{\theta}_i$，$\boldsymbol{\theta}_i$ 可能是多维的. 从上一小节中可以看出，各种判别法则的确定都与 $f_i(\boldsymbol{x},\boldsymbol{\theta}_i)$的取值有关. 而由于 $f_i(\boldsymbol{x},\boldsymbol{\theta}_i)$中有未知参数 $\boldsymbol{\theta}_i$，因此判别法则无法确定. 为克服这一困难就要求我们事先手头有一批数据，用这批数据去估计未知参数 $\boldsymbol{\theta}_i$，$i=1,\cdots,m$.

设

$$\underset{n\times p}{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_m \end{pmatrix}.$$

其中 $\boldsymbol{X}_i(n_i\times p)$为取自子体 Π_i 的样本，$\sum_1^m n_i = n$. 这批样本 $\boldsymbol{X}$ 常

称为**训练样本**(training sample). 由于已知 $\boldsymbol{X}_i$ 是取自 Π_i 的样本，于是便可以用熟知的参数估计方法通过 $\boldsymbol{X}_i$ 去估计 $f_i(\boldsymbol{x},\boldsymbol{\theta}_i)$ 中的未知参数 $\boldsymbol{\theta}_i$，记相应的估计值为 $\hat{\boldsymbol{\theta}}_i$，于是可以用 $f_i(\boldsymbol{x},\hat{\boldsymbol{\theta}}_i)$ 来代替 $f_i(\boldsymbol{x},\boldsymbol{\theta}_i)$ 作为子体 Π_i 的分布，上一小节介绍的各种判别法则便可以使用. 特别当 X 是随机地取自总体 $\mathscr{P}$ 时，当 n 较大时还可以用 $\hat{\pi}_i=\frac{n_i}{n}$ 去估计子体 Π_i 在总体 $\mathscr{P}$ 中的比例 $\pi_i, i=1,\cdots,m$.

例 18.5.8 设 $m=2, \Pi_1=N_p(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}), \Pi_2=N_p(\boldsymbol{\mu}_2,\boldsymbol{\Sigma})$，即两子体有不同的均值和相同的方差阵，$\boldsymbol{\mu}_1,\boldsymbol{\mu}_2$ 和 $\boldsymbol{\Sigma}$ 均未知. $\boldsymbol{X}_i(n_i\times p)$ 为取自 Π_i 的训练样本，$i=1,2$. 用样本均值 $\bar{\boldsymbol{x}}_i$ 去估计子体均值 $\boldsymbol{\mu}_i$，$i=1,2$；用样本合方差阵 $\boldsymbol{S}_p=\frac{n_1\boldsymbol{S}_1+n_2\boldsymbol{S}_2}{n_1+n_2-2}$ 去估计方差阵 $\boldsymbol{\Sigma}$. 根据(18.51)得到贝叶斯判别法则为

$$
\begin{aligned}
R_1 &= \left\{\boldsymbol{x}:\left(\boldsymbol{x}-\frac{\bar{\boldsymbol{x}}_1+\bar{\boldsymbol{x}}_2}{2}\right)^{\mathrm{T}}\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2)>\log\frac{\pi_1}{\pi_2}\right\},\\
R_2 &= R^p-R_1.
\end{aligned}
\tag{18.59}
$$

在上式中若 π_1,π_2 未知，还可用$\frac{n_1}{n}$和$\frac{n_2}{n}$来代替或者根据极大似然法则(18.52)有

$$
\begin{aligned}
R_1 &= \left\{\boldsymbol{x}:\left(\boldsymbol{x}-\frac{\bar{\boldsymbol{x}}_1+\bar{\boldsymbol{x}}_2}{2}\right)^{\mathrm{T}}\boldsymbol{S}_p^{-1}(\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2)>0\right\},\\
R_2 &= R^p-R_1.
\end{aligned}
\tag{18.60}
$$

例 18.5.9 设 $m=2, \Pi_1=N_p(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1), \Pi_2=N_p(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$，即两子体有不同的均值和不同的方差阵，$\boldsymbol{\mu}_1,\boldsymbol{\mu}_2$ 和 $\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2$ 均未知，$\boldsymbol{X}_i(n_i\times p)$ 为取自 Π_i 的训练样本，$i=1,2$. 这时贝叶斯判别有形式

$$
R_1=\left\{\boldsymbol{x}:\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}+c>\log\frac{\pi_2}{\pi_1}\right\},\quad R_2=R^p-R_1,\tag{18.61}
$$

其中 $\boldsymbol{A}=-\frac{1}{2}(\boldsymbol{\Sigma}_1^{-1}-\boldsymbol{\Sigma}_2^{-1}), \boldsymbol{b}=(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2)$,

$$
c=\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|}-\frac{1}{2}(\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2),
$$

在上式中用 $\bar{x}_i$ 代替 μ_i,S_i 代替 Σ_i,$i=1,2$. 即可得到确定的判别法则. 当 π_1,π_2 未知时,可用 $\log\frac{n_2}{n_1}$或 0 来代替 $\log\frac{\pi_2}{\pi_1}$.

注 在上面两例中,为保持判别法则的"稳健性",μ_i,Σ(或 Σ_i)的估计可用适当的稳健估计(见 18.2.2).

例 18.5.10 设 $m=2$,$\Pi_i=\begin{pmatrix}x_{i1}\\x_{i2}\end{pmatrix}$,$i=1,2$,其中 x_{i1}与 x_{i2}为独立的伯努利(Bernoulli)变量,即 $P(x_{ij}=1)=p_{ij}=1-P(x_{ij}=0)$,$j=1,2$,$i=1,2$. 根据 18.5.3 可以推出

$$R_1=\left\{\beta_0+\beta_1x_1+\beta_2x_2\geqslant\log\frac{\pi_1}{\pi_2}\right\},\qquad(18.62)$$

其中 $\beta_0=\log\frac{(1-p_{11})(1-p_{12})}{(1-p_{21})(1-p_{22})}$,$\beta_1=\log\frac{p_{11}(1-p_{21})}{(1-p_{11})p_{21}}$,$\beta_2=\log\frac{p_{12}(1-p_{22})}{(1-p_{12})p_{22}}$. 当存在训练样本时可以用$\frac{n_{ij}}{n_i}$去估计 p_{ij},其中 n_i 为第 i 个子体的样本数,n_{ij}为第 i 个子体样本中 j 变量取 1 的个数. 当然,也可以用 $\log\frac{n_2}{n_1}$或零去估计 $\log\frac{\pi_2}{\pi_1}$.

18.5.4 分布完全未知时的判别法则

当子体 Π_i 的分布 $f_i(\boldsymbol{x})$完全未知时,上一小节的方法完全不能用. 为此需要从训练样本本身的结构出发直接构造判别法则. 下面介绍几种常用方法.

1. 距离判别法

假定有两组训练样本,分别来自两个不同的子体. 现在又有一个观测值 $\boldsymbol{x}$,要判断 $\boldsymbol{x}$ 是取自哪个子体. 一个直观的想法是定出两个子体的中心,例如可以用样本均值 $\bar{\boldsymbol{x}}_i$ 来作为子体 Π_i 的中心,$i=1,2$. 然后看 $\boldsymbol{x}$ 到两个子体中心的距离,如果 $\boldsymbol{x}$ 距 Π_1 中心比 $\boldsymbol{x}$ 距 Π_2 中心近,则判定 $\boldsymbol{x}$ 属于 Π_1,反之则判 $\boldsymbol{x}$ 属于 Π_2. 这就是距离判别法的基本思想,问题是使用什么样的距离函数?

从图 18.1 中可以看出用通常的欧氏距离是不行的. 在图中小圆圈表示第一组训练样本;"×"号表示另一组训练样本. 分别用⊙和⊗表示它们的中心. 观测值 $\boldsymbol{x}$ 到两个中心有相同的欧氏距离. 但直观上看 $\boldsymbol{x}$ 应属于 Π_1 而不属于 Π_2. 其原因是 Π_1 的数据散布度较大,而 Π_2 的数据则较密集. 因此我们在定义距离函数时要考虑这个问题.

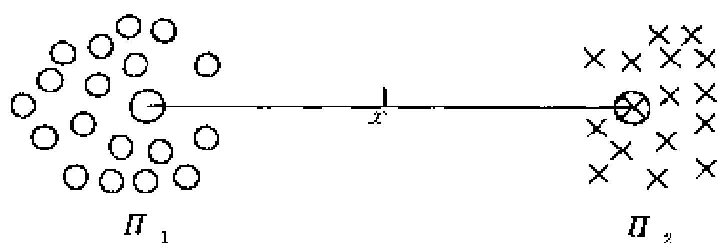


图 18.1

定义 18.5.11 设 R^p 中一个分布有方差矩阵 $\boldsymbol{\Sigma}$, $|\boldsymbol{\Sigma}|>0$, 则在这方差阵下两点 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的马哈拉诺比斯(Mahalanobis)**距离**为

$$d_M(\boldsymbol{x},\boldsymbol{y}) = [(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{y})]^{\frac{1}{2}},$$

简称为**马氏距离**.

容易验证,马氏距离满足一般距离函数的三条基本性质,即

(1) 非负性 $d_M(\boldsymbol{x},\boldsymbol{y})\geqslant 0$; $d_M(\boldsymbol{x},\boldsymbol{y})=0$ 当且仅当 $\boldsymbol{x}=\boldsymbol{y}$.

(2) 对称性 $d_M(\boldsymbol{x},\boldsymbol{y})=d_M(\boldsymbol{x},\boldsymbol{z})$.

(3) 三角不等式 $d_M(\boldsymbol{x},\boldsymbol{y})\leqslant d_M(\boldsymbol{x},\boldsymbol{z})+d_M(\boldsymbol{z},\boldsymbol{y})$.

不仅如此,马氏距离还有一些重要的性质. 我们给出的定义为

定义 18.5.12 马氏模 $\|\boldsymbol{x}\|_M=d_M(\boldsymbol{x},\boldsymbol{0})$.

首先马氏模有刻度不变性. 设 R^p 中一分布有方差阵 Σ, 对 R^p 作一变换 $\boldsymbol{x}\to\boldsymbol{x}^*=\mathrm{diag}(k_1,k_2,\cdots,k_p)\boldsymbol{x}=D\boldsymbol{x}$, 这是一个刻度变换, 相当于把第 i 个分量的刻度单位取为原单位的 $1/k_i$, $i=1,\cdots,p$, 在此变换下方差阵变为 $\boldsymbol{\Sigma}^*=\boldsymbol{D\Sigma D}$. 我们用 $\|\cdot\|_M$ 表示原马氏模

而用 $\|\cdot\|_M^*$ 表示变换后的马氏模. 则容易证明,

$$\|\boldsymbol{x}\|_M = \|\boldsymbol{x}^*\|_M^*, \text{对任意 } x \in R^p. \tag{18.63}$$

特别有

$$\|k\boldsymbol{x}\|_M = k\|\boldsymbol{x}\|_M, \text{对任意 } k > 0. \tag{18.64}$$

其次,马氏模的大小不仅与点 $\boldsymbol{x}$ 的(欧氏)长度有关,也与其方向有关.

定理 18.5.13 设 $\boldsymbol{\Sigma}$ 的特征根为 $\lambda_1 > \lambda_2 > \cdots > \lambda_p$,相应的单位特征向量为 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_p$. 则

$$\|\boldsymbol{\gamma}_1\|_M = \frac{1}{\lambda_1} = \min_{x^{\mathrm{T}}x=1} \|\boldsymbol{x}\|_M,$$

$$\|\boldsymbol{\gamma}_k\|_M = \frac{1}{\lambda_k} = \min_{\substack{x^{\mathrm{T}}x=1 \\ x^{\mathrm{T}}\gamma_j=0, j=1,\cdots,k-1}} \|\boldsymbol{x}\|_M, k = 2, \cdots, p.$$

熟知,$\boldsymbol{\gamma}_1$ 是分布散布度最大的方向,因此,沿此方向上马氏距离最短. $\boldsymbol{\gamma}_2$ 是在和 $\boldsymbol{\gamma}_1$ 正交的方向上分布散布度最大的方向,因此沿此方向上的马氏距离在所有与 $\boldsymbol{\gamma}_1$ 正交的方向上是最短的,……. 而在 $\boldsymbol{\gamma}_p$ 方向上马氏距离最长.

最后,我们来解答图 18.1 所提出的问题.

定理 18.5.14 设 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$ 为两个方差阵,且 $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2 > 0$(即 $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$ 为正定阵). 用 $\|\cdot\|_M^{(i)}$ 定义由 $\boldsymbol{\Sigma}_i$ 决定的马氏模. 则对任意的 x 有

$$\|\boldsymbol{x}\|_M^{(1)} < \|\boldsymbol{x}\|_M^{(2)}.$$

上述定理说明,协方差愈"大",则马氏模愈小,这就回答了在图 18.1 中为什么 $\boldsymbol{x}$ 应属于 Π_1 而不属于 Π_2 的道理,与直观的感觉是一致的. 下面给出距离判别法的正式定义.

定义 18.5.15 设 m 个子体 $\Pi_1, \cdots, \Pi_m$ 分别有均值 $\mu_1, \cdots, \mu_m$ 和方差阵 $\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_m$. 对任一观测值 $\boldsymbol{x}$,定义 $\boldsymbol{x}$ 到子体 Π_i 的距离为

$$d(\boldsymbol{x}, \Pi_i) = d_M^{(i)}(\boldsymbol{x}, \mu_i) = [(\boldsymbol{x} - \mu_i)^{\mathrm{T}} \textstyle\sum_i^{-1} (\boldsymbol{x} - \mu_i)]^{\frac{1}{2}},$$

$$i = 1,\cdots,m.$$

则距离判别法则为

$$R_i = \{\boldsymbol{x}: d(\boldsymbol{x},\Pi_i) = \min_{1\leqslant j\leqslant m} d(\boldsymbol{x},\Pi_j)\}, i = 1,\cdots,m.$$

当 μ_i 和 $\boldsymbol{\Sigma}_i$ 均未知而存在训练样本 $\boldsymbol{X}_i$ 时，可用估计值 $\hat{\mu}_i$ 和 $\hat{\boldsymbol{\Sigma}}_i$ 来定义马氏距离，从而确定距离判别准则.

例 18.5.16 设 $m=2$，Π_i 有均值 μ_i 和方差阵 $\boldsymbol{\Sigma}_i$，$i=1,2$. 设 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$. 于是 $d(\boldsymbol{x},\Pi_i)=(\boldsymbol{x}-\mu_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu_i)$，$i=1,2$. 因此距离判别有形式

$$R_1 = \{d(\boldsymbol{x},\Pi_1) \leqslant d(\boldsymbol{x},\Pi_2)\},$$
$$R_2 = \{d(\boldsymbol{x},\Pi_1) > d(\boldsymbol{x},\Pi_2)\}.$$

经计算容易得出

$$R_1 = \left\{\left(\boldsymbol{x} - \frac{\mu_1 + \mu_2}{2}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mu_1 - \mu_2) \geqslant 0\right\},$$
$$R_2 = R^p - R_1. \tag{18.65}$$

这与正态等方差阵条件下 $\pi_1=\pi_2$ 时的贝叶斯判别法则一致（见(18.55)）.

例 18.5.17 在上例中若 $\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_2$. 则距离判别与两正态子体不等方差阵条件下 $\pi_1=\pi_2$ 时的贝叶斯判别法则类似（见(18.61)），在那里 c 改为 $-\frac{1}{2}(\mu_1^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\mu_1-\mu_2^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}\mu_2)$，$\log\frac{\pi_2}{\pi_1}$ 用 0 代替即可.

注 在用距离判别时，由训练样本估计出来的均值和方差阵会发生不稳健性的问题. 但是在使用稳健估计时会遇到一种情形，即估计的方差阵会产生不定性. 因此会发生这样的怪现象，即对某些观测值其马氏距离平方会取负值！但幸运的是，这种情形（即距离平方为负）在实际中很少发生. 因此为了保证判别的稳健性，冒这一点小小的风险还是值得的.

在使用距离判别时会遇到一个比较严重的问题，即方差阵 $\boldsymbol{\Sigma}$

退化的情形，这时相应的方差阵的估计有一些特征值非常接近零（与其它特征值比较而言）. 这时分布本质上是退化的，即集中在一个 k 维子空间上，$k<p$，而 $\boldsymbol{\Sigma}$ 的估计 $\hat{\boldsymbol{\Sigma}}$ 的后 $p-k$ 个特征根可以看成是由观测误差所造成的（参见下一节主成份分析）. 在这种情形下，即使 $\hat{\boldsymbol{\Sigma}}^{-1}$ 存在，由 $\hat{\boldsymbol{\Sigma}}^{-1}$ 所定义的马氏距离有一个严重的缺点，即在 $\hat{\boldsymbol{\Sigma}}^{-1}$ 的后几个特征向量的方向上，向量长度的微小变化都会引起马氏距离的很大改变. 在判别中就会因微小的观测误差而引起严重误判. 为克服这一困难，我们可以适当选取 $\hat{\boldsymbol{\Sigma}}$ 的前 k 个特征向量（主成份）$\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_k$（具体选法见本章 6.2.3）. 记

$$\widetilde{\boldsymbol{\Gamma}}=(\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_k),\widetilde{\boldsymbol{D}}=\mathrm{diag}(\lambda_1,\cdots,\lambda_k),$$

$$\hat{\boldsymbol{\Sigma}}^{+}=\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{D}}^{-1}\widetilde{\boldsymbol{\Gamma}}^{\mathrm{T}}. \tag{18.66}$$

对每一观测值 $\boldsymbol{x}$，定义其到分布中心的马氏距离为

$$d_M^{+}(\boldsymbol{x},\Pi)=(\boldsymbol{x}-\hat{\mu})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{+}(\boldsymbol{x}-\hat{\mu}). \tag{18.67}$$

这样做的结果实质上是用向量在前 k 个特征向量所生成子空间的投影来定义马氏距离. 在距离判别中当 p 较大时，用 $d_M^{+}(\boldsymbol{x},\Pi)$ 来代替 $d_M(\boldsymbol{x},\Pi)$ 其效果更好.

2. 费希尔判别法

在前面介绍的一些判别法中常出现观测值 $\boldsymbol{x}$ 的某种线性函数，根据线性函数的大小来判别观测值属于哪个子体. 费希尔(Fisher)借助方差分析的思想来导出线性判别函数，其思想是这样：找一个方向，在这个方向上要使各子体内部尽可能“密集”；而子体之间要尽可能“分开”. 然后将观测值沿这方向投影，根据投影值的大小来进行判别.

用 p 维随机变量 $\boldsymbol{y}_i$ 来表示子体 $\Pi_i,i=1,\cdots,m$，对任一给定的方向 $\boldsymbol{u}$，$\boldsymbol{y}_i$ 在 $\boldsymbol{u}$ 方向的投影为 $\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{u},i=1,\cdots,m$. 记

$$e_i=E(\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{u})=\mu_i^{\mathrm{T}}\boldsymbol{u},i=1,\cdots,m,\bar{e}=\frac{1}{m}\sum_{i=1}^{m}e_i, \tag{18.68}$$

$$\sigma_i^2 = V(\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{u}, i = 1, \cdots, m.$$

费希尔用“组间离差”

$$B(\boldsymbol{u}) = \sum_{i=1}^{m}(e_i - \bar{e})^2 = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{MHM}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{Bu} \quad (18.69)$$

来刻划在 $\boldsymbol{u}$ 方向子体之间分开的程度，其中

$$\boldsymbol{M} = (\mu_1, \cdots, \mu_m), \boldsymbol{H} = \boldsymbol{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}. \quad (18.70)$$

用“组内离差”

$$E(\boldsymbol{u}) = \sum_{i=1}^{m}\sigma_i^2 = \boldsymbol{u}^{\mathrm{T}}\left(\sum_{1}^{m}\boldsymbol{\Sigma}_i\right)\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{Eu} \quad (18.71)$$

来刻划在 $\boldsymbol{u}$ 方向子体内部的聚集程度. 费希尔认为应取 $\boldsymbol{u}$ 使得 $B(\boldsymbol{u})/E(\boldsymbol{u})$ 达到最大.

定义 18.5.18 设 p 维向量 $\boldsymbol{u}_1$ 满足：

$$\frac{\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Bu}_1}{\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Eu}_1} = \max_{u}\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Bu}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Eu}},$$

则对任何观测值 $\boldsymbol{x}$，$\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{x}$ 称为**费希尔线性判别函数**.

定理 18.5.19 满足上述定义的 $\boldsymbol{u}_1$ 为 $\boldsymbol{E}^{-1}\boldsymbol{B}$ 的最大特征根所对应的特征向量.

注 在实际求解 $\boldsymbol{u}_1$ 时，先求 $\boldsymbol{E}^{-\frac{1}{2}}\boldsymbol{BE}^{-\frac{1}{2}}$（对称阵）的最大特征根，找到特征向量 $\boldsymbol{v}_1$ 时，令 $\boldsymbol{u}_1 = \boldsymbol{E}^{-\frac{1}{2}}\boldsymbol{v}_1$ 即为所求.

定理 18.5.19 只是给出了确定线性判别函数的投影方向 $\boldsymbol{u}_1$，而并没有给出具体的判别法则. 由于 $\boldsymbol{u}_1\boldsymbol{x}$ 是一维的，我们可以用一维空间中的马氏距离来判别，得到结果如下.

将 $e_i, i=1,\cdots,m$ 从小到大排列为 $e_{(1)} \leqslant e_{(2)} \leqslant \cdots \leqslant e_{(m)}$. 取 $e_{(i)}$ 和 $e_{(i+1)}$ 的加权中点 $m_i = \dfrac{\sigma_{(i)}^2 e_{(i)} + \sigma_{(i+1)}^2 e_{(i+1)}}{\sigma_{(i)}^2 + \sigma_{(i+1)}^2}, i=1,\cdots,m-1$. 于是得到如下的

方法 18.5.20 判别法则

当 $\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{x} \in (m_{i-1}, m_i]$ 时，判定 $\boldsymbol{x}$ 属于 $\Pi_{(i)}, i=1,\cdots,m, m_0=$

$-\infty, m_{m+1}=+\infty$.

特别当 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\cdots=\boldsymbol{\Sigma}_m$ 时，$m_i=\dfrac{e_{(i)}+e_{(i+1)}}{2}, i=1,\cdots,m$.

上面讲的是用一个线性判别函数来进行判别的情形. 有时用一个线性判别函数判别还不够(特别是靠近边界点的那些地方)，需要同时用几个线性判别函数. 我们可取 $\boldsymbol{E}^{-1}\boldsymbol{B}$ 的前 k 个特征根所对应的特征向量 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k$；$\boldsymbol{u}_i$ 称为第 i 个正则方向，$i=1,\cdots,k$. 用 $(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{u}_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{u}_k)^{\mathrm{T}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}$ 去进行判别，其中 $\boldsymbol{U}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_k)$. 这相当于将原来的 p 维向量降维到 k 维，对应于子体 Π_i 的均值为 $\tilde{\boldsymbol{\mu}}_i=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\mu}_i$，方差阵 $\tilde{\boldsymbol{\Sigma}}_i=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{U}, i=1,\cdots,m$. 于是可以用距离判别法则对 $\boldsymbol{z}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}$ 进行判别.

在子体分布完全未知的情况下，除了上面介绍的距离判别法和费希尔判别法之外，还有一些非参数判别法，如核(kernal)判别和近邻(nearest neighbour)判别.

18.5.5 误判概率

在判别分析中误判概率的计算与估计是一个重要的问题. 在判别分析中误判概率的大小是由两个方面的原因决定的. 其一是由所要判别的那些子体决定的，如果子体之间是很难区分的，那么即使采用理论上最好的判别法误判概率仍然可能很大. 其二是由所使用的方法决定的. 对同一判别问题使用不同的判别法会产生不同的误判概率. 因此，通过计算和估计误判概率可以使我们在不同的判别法中选择一个较为满意的(误判概率较小的)判别法. 另一方面，如果采用任何判别法，误判概率都不能显著变小，那么就可以认为这个判别问题本身“很坏”，不能从中推出任何较为有意义的结论.

首先我们介绍在各子体的分布已知的条件下计算误判概率的方法. 实际上我们已经在本节的开头给出了计算公式即(18.52).

根据这个公式，首先要计算概率 $P_{err}(j|i)=\int_{R_j} f_i(\boldsymbol{x})\mathrm{d}x$，然后再计算总平均误判概率 P_{err}.

例 18.5.21 设有两个子体，$\Pi_1 \sim N_p(\mu_1,\boldsymbol{\Sigma})$，$\Pi_2 \sim N_p(\mu_2,\boldsymbol{\Sigma})$. 这时贝叶斯判别，距离判别和费希尔判别有共同的线性判别函数.

$$h(\boldsymbol{x})=\left(\boldsymbol{x}-\frac{\mu_1+\mu_2}{2}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mu_1-\mu_2). \qquad (18.72)$$

而判别域为

$$R_1=\{\boldsymbol{x}: h(\boldsymbol{x})\geqslant a\},\ R_2=\{x: h(\boldsymbol{x})<a\}. \qquad (18.73)$$

其中，在距离判别下 $a=0$；在贝叶斯判别下，$a=\log(\pi_1/\pi_2)$；在费希尔判别下 a 可以取 0 或别的数. 此时可以证明：

$$h(\boldsymbol{x})\sim N\left(\frac{\Delta}{2},\Delta\right),\text{当 } \boldsymbol{x}\in\Pi_1, \qquad (18.74)$$

$$h(\boldsymbol{x})\sim N\left(-\frac{\Delta}{2},\Delta\right),\text{当 } \boldsymbol{x}\in\Pi_2,$$

其中

$$\Delta=(\mu_1-\mu_2)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mu_1-\mu_2). \qquad (18.75)$$

因此

$$P_{err}(2|1)=P(h(\boldsymbol{x})<a|x\in\Pi_1)=\Phi\left(\frac{a-\frac{\Delta}{2}}{\sqrt{\Delta}}\right), \qquad (18.76)$$

$$P_{err}(1|2)=P(h(\boldsymbol{x})\geqslant a|x\in\Pi_2)=1-\Phi\left(\frac{a+\frac{\Delta}{2}}{\sqrt{\Delta}}\right). \qquad (18.77)$$

而总平均误判概率为

$$P_{err}=\pi_1\Phi\left(\frac{a-\frac{\Delta}{2}}{\sqrt{\Delta}}\right)+\pi_2\left(1-\Phi\left(\frac{a+\frac{\Delta}{2}}{\sqrt{\Delta}}\right)\right), \qquad (18.78)$$

其中 $\Phi(\cdot)$ 为 $N(0,1)$ 的分布函数，可查表求得.

对 minimax 判别法则，应满足

$$P(2|1)=P(1|2),\text{即}\ \Phi\left(\frac{a-\frac{\Delta}{2}}{\sqrt{\Delta}}\right)=1-\Phi\left(\frac{a+\frac{\Delta}{2}}{\sqrt{\Delta}}\right). \tag{18.79}$$

据此可以定出 a 应满足

$$\frac{a-\frac{\Delta}{2}}{\sqrt{\Delta}}=\frac{-\left(a+\frac{\Delta}{2}\right)}{\sqrt{\Delta}},\text{即}\ a=0. \tag{18.80}$$

由(18.78)式可见，当 Δ 大时，P_{err} 小，反之当 Δ 小时 P_{err} 大，从距离的观点看，Δ 表示两子体中心距离的平方. 因此上述事实实际上说明：当两子体中心距离越大，则越容易判别，反之就不易判别.

当分布含有未知参数时，误判概率也是参数的函数，因而是不确定的. 这时如果手头上有训练样本的话就可以用它去估计参数值. 用这些估计值作真值，按照理论上给出的公式去计算误判概率，实际上得到的是误判概率的估计值. 这方法最简便易行，但在样本较小时，所得的误判概率的估计不够好.

下面介绍几种完全不依赖于分布的误判概率估计法.

方法 18.5.22　回报法　假定有一批训练样本 $x_{ij}\in\Pi_i$，$j=1,\cdots,n_i$，$i=1,\cdots,m$，$\Sigma n_i=n$. 依照某种判别法则用这批样本得出判别域为 $R_1,\cdots,R_m$. 用这组判别域返回来对训练样本逐个进行判别. 记

$n_{err}(j|i)$＝第 i 个子体的样本被误判为

属于 j 子体的个数，$j\neq i$，$i=1,\cdots,m$.

$$n_{err}=\sum_{i=1}^{m}\sum_{j\neq i}n_{err}(j|i). \tag{18.81}$$

于是可以估计条件误判概率为

$$\hat{P}_{err}(j|i)=\frac{n_{err}(j|i)}{n_i}. \tag{18.82}$$

当先验概率 $\pi_1,\cdots,\pi_m$ 已知时，可用

$$\hat{P}_{err} = \sum_{i=1}^{m} \pi_i \sum_{j \neq i} \hat{P}_{err}(j|i) \tag{18.83}$$

去估计总的误判概率.

当 $\pi_1, \cdots \pi_m$ 未知时，可用 $\frac{n_i}{n}$ 去估计 π_i. 于是有

$$\hat{P}_{err} = n_{err}/n. \tag{18.84}$$

理论上和实践上都证明，用回报法所得到的误判概率比真实的误判概率偏小，这是回报法的一个缺点.

方法 18.5.23　刀切(jackknife)法　这是对回报法的一个修正. 记用某种判别法则从 n 个训练样本得到的判别域为 $R_1^{(n)}, \cdots, R_m^{(n)}$. 现在要返回去判别 $\boldsymbol{x}_i$ 时，不用 $\{R_j^{(n)}\}$，而是从原来的 n 个样本中去掉 $\boldsymbol{x}_i$ 后剩下的 $(n-1)$ 个样本去构造判别域为 $R_{1i}^{(n-1)}, \cdots R_{mi}^{(n-1)}$，用这一组判别域去判别 $\boldsymbol{x}_i$. $n_{err}(j|i), n_{err}, \hat{P}_{err}(j|i)$ 及 $\hat{P}_{err}$ 仍如 18.5.22 中所定义，这样得到的误判概率估计比用回报法更接近真值. 当然刀切法的计算量较大，可以使用计算机来解决.

方法 18.5.24　自助(bootstrap)法　设用回报法得到 $n_{err}(j|i)$ 及 $\hat{P}_{err}(j|i)$. 由于 $\hat{P}_{err}(j|i)$ 是 $P_{err}(j|i)$ 的有偏估计，记 $d_{ij} = P_{err}(j|i) - \hat{P}_{err}(j|i)$，现要对偏差 d_{ij} 进行估计. 用训练样本 $x_{i_1}, \cdots, x_{in_i}$ 代替子体 $\Pi_i, i=1, \cdots, m$，从中进行有放回抽样 n_i 次，得到新的样本 $x_{i_1}^*, \cdots, x_{in_i}^*, i=1, \cdots, m$. 由于是有放回抽样，$\{x_{ij}^*\}$ 可看成是独立的取自子体 $\hat{\Pi}_i = \{x_{ij}, j=1, \cdots, n_i\}$ 的随机样本. 用这一组新的样本构造新的判别域 $R_1^*, \cdots, R_m^*$，然后由这组判别域出发对 $\{x_{ij}\}$ 和 $\{x_{ij}^*\}$ 进行判别，分别得到 $n_{err}^*(j|i)$ 和 $n_{err}^{**}(j|i)$. 记

$$d_{ij}^* = (n_{err}^*(j|i) - n_{err}^{**}(j|i))/n_i, \tag{18.85}$$

d_{ij}^* 为一次重新抽样中 d_{ij} 的估计. 将这种重新抽样进行 K 次，K 可以很大($K=100$ 或 500)，第 k 次得到的 $d_{ij}^*(k)$ 按(18.83)所定义. 于是对 d_{ij} 的估计可用

$$\hat{d}_{ij} = \frac{1}{K} \sum_{k=1}^{K} d_{ij}^*(k) \tag{18.86}$$

给出，于是 $P_{err}(j|i)$的自助估计为

$$\hat{P}_b(j|i) = \hat{P}_{err}(j|i) + \hat{d}_{ij}, \quad i \neq j. \tag{18.87}$$

总平均误判概率的自助估计为

$$\hat{P}_{err \cdot b} = \sum_{i=1}^{m} \pi_i \sum_{j \neq i} \hat{P}_b(j|i). \tag{18.88}$$

这里，估计偏差是从训练样本中进行再抽样得到的. 这种再抽样在计算机上进行只是个时间问题.

大量的模拟计算证明，用刀切法和自助法所得出的误判概率估计是比较好的.

18.5.6 选择变量

在判别分析中如同在回归分析中一样，存在着选择变量的问题. 在判别分析中考虑选择变量的问题就是要研究所关心的这个(或这些)变量对区分各个子体是否有好的效果. 这里有几种情形会发生，我们以两子体两变量为例来作一简单说明.

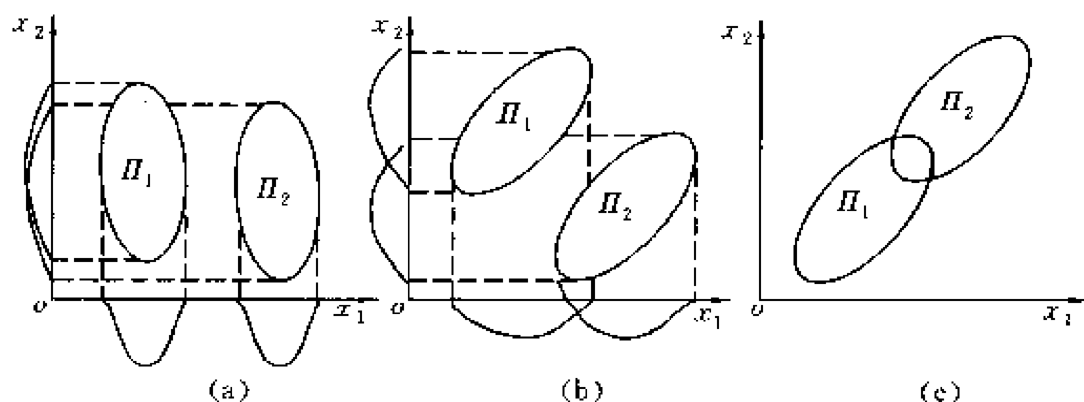


图 18.2

在图 18.2 中有三种情形. (a)显示变量 x_1 对区分子体 Π_1, Π_2 有明显效果而 x_2 则几乎不起作用. (b)显示两变量各自单独对判别 Π_1, Π_2 都效果不太好，而联合起来则效果较好. (c)显示仅用 x_1 或仅用 x_2 判别与将二者联合起来进行判别效果差不多，保留一个就行了. 当变量个数多的时候情况就更复杂了.

为了解决选择哪些变量来进入判别法则的问题，我们首先要对所研究的变量(比如说 p 个变量中的前 r 个，$r\leqslant p$)定义一个数值指标 $\Delta_{1,2,\cdots,r}$，这指标值越大则反映这些变量区分子体的能力越强. 然后对这指标定一个合理的阈值 T，当 $\Delta>T$ 时说明这些变量的判别力是好的. 否则就不好. 以下我们介绍几个用于确定变量判别能力的数量指标，它们都可以称为辨别率. 为简化问题起见总假设子体是两个：Π_1,Π_2，记所考虑的全部变量为 $x_1,x_2,\cdots,x_p$.

定义 18.5.25 设子体 Π_i 关于 p 维向量 $\boldsymbol{x}=(x_1,\cdots,x_p)^{\mathrm{T}}$ 有均值向量 $\boldsymbol{\mu}_i$，方差阵 $\boldsymbol{\Sigma}_i$，$i=1,2$，定义 $\boldsymbol{x}$ 的距离分辨率指标为

$$\Delta=\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}}(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2).$$

距离分辨率是分别在两子体的协方差阵下计算两子体中心(均值)的马氏距离平方再求和. 特别当 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ 时，有

$$\Delta=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2). \tag{18.89}$$

18.5.25 所给出的只是形式上的分辨率. 因为 $\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}_i$ 均未知. 当存在训练样本 $\boldsymbol{x}_{ij}$，$j=1,\cdots,n_i$，$i=1,2$ 时，用样本均值 $\bar{\boldsymbol{x}}_i$ 和样本方差阵 $\hat{\boldsymbol{\Sigma}}_i$ 去代替 $\boldsymbol{\mu}_i$ 及 $\boldsymbol{\Sigma}_i$，便得到实在的分辨率.

现在考虑 p 个变量中的一部分(不妨设为前 r 个)的分辨率，我们用下面的记号来记向量和矩阵的分块形式.

$$\underset{(p\times 1)}{\boldsymbol{a}}=\begin{pmatrix}\boldsymbol{a}^{(1)}\\ \boldsymbol{a}^{(2)}\end{pmatrix}\begin{matrix}r\\ p-r\end{matrix},$$

$$\underset{p\times p}{\boldsymbol{A}}=\begin{pmatrix}\boldsymbol{A}^{11} & \boldsymbol{A}^{12}\\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22}\end{pmatrix}\begin{matrix}r\\ p-r\end{matrix}. \tag{18.90}$$
$$\qquad\quad\; r \qquad p-r$$

于是 x 的前 r 个向量为 $\boldsymbol{x}^{(1)}$，均值为 $\boldsymbol{\mu}_i^{(1)}$，方差阵为 $\boldsymbol{\Sigma}_i^{11}$，当 $\boldsymbol{x}$ 属于 Π_i，$i=1,2$. 记 $\boldsymbol{x}^{(1)}$ 的分辨率为 Δ_r，则有

$$\Delta_r=\frac{1}{2}(\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_2^{(1)})^{\mathrm{T}}\left((\boldsymbol{\Sigma}_1^{11})^{-1}+(\boldsymbol{\Sigma}_2^{11})^{-1}\right)(\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_2^{(1)}). \tag{18.91}$$

当 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ 时，$\boldsymbol{\Sigma}_1^{11}=\boldsymbol{\Sigma}_2^{11}=\boldsymbol{\Sigma}^{11}$，有

$$\Delta_r=(\mu_1^{(1)}-\mu_2^{(1)})^{\mathrm{T}}(\boldsymbol{\Sigma}^{11})^{-1}(\mu_1^{(1)}-\mu_2^{(1)}). \qquad (18.92)$$

可以证明

$$\Delta_p \geqslant \Delta_r. \qquad (18.93)$$

为考察后 $m-r$ 个变量对提高分辨率是否有作用，一个自然的想法是看 $(\Delta_p-\Delta_r)/\Delta_r$ 的大小，在正态等方差的条件下有一个理论上的结果.

定理 18.5.26 设 $\Pi_i\sim N_p(\mu_i,\boldsymbol{\Sigma})$，$i=1,2$. $\boldsymbol{X}_i(n_i\times p)$ 为训练样本，$i=1,2$. 当 Δ_p 和 Δ_r 是用样本均值和样本方差阵计算时有

$$F=\frac{n-p-1}{p-r}\,\frac{\Delta_p-\Delta_r}{c+\Delta_r}\sim F(p-r,n-p-1),$$

其中 $n=n_1+n_2$，$c=n(n-2)/(n_1n_2)$.

这定理给出了一个 F 检验. 但在实际使用时正态假定未必成立，方差阵也不一定相等. 因此在使用上述 F 统计量时不必拘泥于按照严格的 F 检验进行. 通常只要给定一个阈值 F_T（例如 $F_T=10$）即可进行检验了.

在进行向前或向后的逐步判别时，总是考虑 r 个变量与 $r+1$ 个变量的关系. 这时 F 统计量有形式

$$F=(n-r-2)\,\frac{\Delta_{r+1}-\Delta_r}{c+\Delta_r}. \qquad (18.94)$$

因此，当 $F\geqslant F_T$ 时，就在已入选的 r 个变量中再添加第 $r+1$ 个或从已入选的 $r+1$ 个变量中不剔除所检验的那个变量；否则，就不在已入选的 r 个变量中添加检验的另一个变量，或在已入选的 $r+1$ 个变量中剔除所检验的那一个.

距离分辨率指标通常用于分布未知的情形，但是在正态分布同方差阵的条件下，我们有更好的检验法，即利用第 18.3 节中介绍的关于均值检验的 Λ 统计量.

沿用定义 18.5.68 中的记号，要考虑全部 p 个变量是否对判

别起作用，实质上是检验假设

$$H: \mu_1 = \mu_2. \tag{18.95}$$

当 H 被接受时，说明这 p 个变量无法区分 Π_1, Π_2；否则可以区分. 根据定理 18.3.11，检验统计量可取

$$\Lambda_{(p)} = \frac{|\boldsymbol{W}|}{|\boldsymbol{T}|} = \frac{|\boldsymbol{W}|}{|\boldsymbol{B}+\boldsymbol{W}|}, \tag{18.96}$$

其中

$$\boldsymbol{W} = \sum_1^2 n_i \boldsymbol{S}_i = \sum_1^2 \boldsymbol{X}_i^{\mathrm{T}}\left(\boldsymbol{I}_{n_i} - \frac{1}{n_i}\mathbf{1}_{n_i}\mathbf{1}_{n_i}^{\mathrm{T}}\right)\boldsymbol{X}_i, \tag{18.97}$$

$$\boldsymbol{T} = n\boldsymbol{S} = \boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\boldsymbol{X},$$

$$\boldsymbol{B} = \boldsymbol{T} - \boldsymbol{W} = \sum_1^2 n_i(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})^{\mathrm{T}}.$$

$\boldsymbol{W}, \boldsymbol{T}$ 和 $\boldsymbol{B}$ 分别称为组内离差阵，全离差阵和组间离差阵. 根据定理 18.3.12 有

$$\Lambda_{(p)} \sim \Lambda(p, n-2, 1). \tag{18.98}$$

对给定的水平 α，记 $\Lambda_\alpha(p, n-2, 1)$ 为 $\Lambda(p, n-2, 1)$ 的下 α 分位点，即

$$P(\Lambda(p, n-2, 1) < \Lambda_\alpha(p, n-2, 1)) = \alpha. \tag{18.99}$$

则假设(18.95)的检验法则为

$$\text{否定域} \qquad W = \{\Lambda_{(p)} < \Lambda_\alpha(p, n-2, 1)\}. \tag{18.100}$$

假定要考虑 p 个变量中的 r 个是否对判别起作用，不妨假定是前 r 个. 沿用(18.90)所规定的记号，相当于检验假设

$$H_1: \mu_1^{(1)} = \mu_2^{(1)}. \tag{18.101}$$

相应的 Λ 统计量为

$$\Lambda_{(r)} = \frac{|\boldsymbol{W}^{11}|}{|\boldsymbol{T}^{11}|} = \frac{|\boldsymbol{W}^{11}|}{|\boldsymbol{B}^{11}+\boldsymbol{W}^{11}|} \sim \Lambda(r, n-2, 1).$$

当此 r 个变量通过了检验，即否定了 H_1 之后，我们进而考虑其余 $p-r$ 个变量是否对判别法则能提供进一步的附加信息. 实际上是

要检验在给定前 r 个变量时，后 $p-r$ 个变量的条件均值是否相同，即检验假定

$$H_{2|1}:\ E_1(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}) = E_2(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}), \tag{18.102}$$

其中 $E_i(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)})$ 表示在子体 Π_i 中给定 $\boldsymbol{x}^{(1)}$ 的条件下 $\boldsymbol{x}^{(2)}$ 的条件期望. 这时可以证明，相应的 Λ 统计量为

$$\Lambda_{(p-r)|(r)} = \frac{\Lambda_{(p)}}{\Lambda_{(r)}} \sim \Lambda(p-r, n-r-2, 1). \tag{18.103}$$

由上式中可看出 Λ 统计量有类似于条件概率公式的性质. 这性质使得逐步判别分析特别方便. 如果从 $\Lambda_{(p-r)|r}$ 出发检验而否定了 $H_{2|1}$，则说明其余 $p-r$ 个变量对判别还能提供信息. 在逐步判别中要入选 $p-r$ 个变量中的一个. 我们记

$$\Lambda_{i_1,\cdots,i_l} \tag{18.104}$$

为由变量 $\boldsymbol{x}_{i_1},\cdots,\boldsymbol{x}_{i_l}$ 所生成的 Λ 统计量，而记

$$\Lambda_{j|i_1,\cdots,i_l},\quad j\neq i_1,\cdots,i_l, \tag{18.105}$$

为当变量 $\boldsymbol{x}_{i_1},\cdots,\boldsymbol{x}_{i_l}$ 已入选后，检验变量 x_j 所提供的附加信息的 Λ 统计量，于是我们 入选 $\boldsymbol{x}_j$，当 $\Lambda_{j|1,\cdots,r}$ 达到最小，对 $j=r+1,\cdots,p$ 由(18.96)可以衍生出一个递推展开公式

$$\begin{aligned}\Lambda_{(p)} &= \Lambda_1\Lambda_{2|1}\Lambda_{3|1,2}\cdots\Lambda_{p|1,\cdots,p-1}\\ &= \Lambda_1\Lambda_{(p-1)|(1)}\\ &= \cdots = \Lambda_{(r)}\Lambda_{(p-r)|(r)}.\end{aligned} \tag{18.106}$$

具体的计算过程限于篇幅便不再详细介绍，请参阅[18].

18.6 主成份分析

18.6.1 主成份的定义

定义 18.6.1 设 $\boldsymbol{x}$ 为 p 维行随机向量，其方差阵为 $\boldsymbol{\Sigma}$. 若 $\boldsymbol{\Sigma}$ 的特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$，相应的标准正交的特征向量为 $\boldsymbol{Y}_1, \boldsymbol{Y}_2$,

$\cdots,\gamma_p$,则称 $y_k=x\gamma_k$ 为 x 的第 k 个**主成份**(principal component),$k=1,2,\cdots,p$.

定理 18.6.2 (1) 第 k 个主成份 $y_k=x\gamma_k$ 有方差 $\lambda_k,k=1,\cdots,p$.

(2) 若 $i\neq j$,则第 i 个主成份 y_i 与第 j 个主成份 y_j 互不相关,即 $\mathrm{cov}(y_i,y_j)=0$.

记 $\boldsymbol{\Sigma}=\Gamma^{\mathrm{T}}\Lambda\Gamma,\Gamma=(\gamma_1,\cdots,\gamma_k),\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p),\boldsymbol{y}=(y_1,\cdots,y_p)$,则有

$$\boldsymbol{y}=\boldsymbol{x}\Gamma,\ \mathrm{v_{ar}}(\boldsymbol{y})=\Lambda. \tag{18.107}$$

所以,主成份分析就是将原随机向量 x 作一正交变换,使得新的随机向量的每一分量都是原向量的一个线性组合,且它们之间两两不相关.

定理 18.6.3 主成份有下述重要特征:

$$V(y_1)=\max_{\|a\|=1}V(xa),$$

$$V(y_2)=\max_{\substack{\|a\|=1\\ a^{\mathrm{T}}\gamma_1=0}}V(xa),$$

$$\cdots\cdots$$

$$V(y_{p-1})=\max_{\substack{\|a\|=1\\ a^{\mathrm{T}}\gamma_1=0,\cdots,a^{\mathrm{T}}\gamma_{p-2}=0}}V(xa),$$

$$V(y_p)=\min_{\|a\|=1}V(xa).$$

由定理 18.6.2 及 18.6.3 可见,第 1 主成份是在 $\|a\|=1$ 的条件下,x 的所有线性组合 xa 中方差最大的,因而刻划了 x 的最主要特征;第 2 主成份是在 $\|a\|=1$ 且与第 1 主成份不相关的约束下,x 的所有线性组合 xa 中方差最大的,因而刻划了 x 的第 2 个主要特征;$\cdots$,第 k 主成份是在 $\|a\|=1$ 且与第 $1,2,\cdots,k-1$ 主成份不相关的约束下,x 的所有线性组合 xa 中方差最大的,因而刻划了 x 的第 k 个主要特征,$\cdots$,等等.

在实际问题中,由于 x 的分布并不知道,因此不能直接按照定

义 18.6.1 来求出 x 的主成份，而必须借助于观测数据.

定义 18.6.4 设 $\boldsymbol{x}$ 的数据阵为 $\boldsymbol{X}_{(n\times p)}$，样本方差阵为 $\boldsymbol{S}=\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I}_n-\frac{1}{n}\boldsymbol{1}_n^{\mathrm{T}}\boldsymbol{1}_n\right)\boldsymbol{X}$. 设 $\boldsymbol{S}$ 的特征根为 $\hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_p$，相应的标准正交特征向量为 $\hat{\boldsymbol{Y}}_1,\cdots,\hat{\boldsymbol{Y}}_p$，则 $\boldsymbol{y}_k=\boldsymbol{X}\hat{\boldsymbol{Y}}_k$ 称为 X 的第 k 个**样本主成份向量**，$k=1,\cdots,p$.

记 $\boldsymbol{Y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_p)$，则 $\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\Gamma}}$，其中 $\hat{\boldsymbol{\Gamma}}=(\hat{\boldsymbol{Y}}_1,\cdots,\hat{\boldsymbol{Y}}_p)$. 由于 $\boldsymbol{S}$ 为 $\boldsymbol{\Sigma}$ 的估计，因此 $\boldsymbol{y}_k$ 相当于第 k 个主成份在 n 次观测中的估计值，其方差估计为 $\hat{\lambda}_k$，$k=1,\cdots,p$.

在深入介绍主成份分析的应用之前，先要指出它的一个主要的缺点，即不具备**尺度不变性**(scale invariance). 在实际问题中，p 维随机变量 $\boldsymbol{x}$ 的各个分量可能是不同性质的量(例如有的是长度，有的是重量，有的是体积等)；或者是同一性质的量但取不同的单位(例如同是长度，有的以米为单位，而有的以厘米为单位). 因此当用不同的尺度来记录 $\boldsymbol{x}=(x_1,\cdots,x_p)$时，相当于将 $\boldsymbol{x}$ 作一变换到 $\boldsymbol{x}^*=\boldsymbol{x}\boldsymbol{D}$，$\boldsymbol{D}=\mathrm{diag}(d_1,d_2,\cdots,d_p)$. 假定 $d_1,\cdots,d_p$ 不全相等，则由 $\boldsymbol{x}^*$ 出发作主成份分析与从 $\boldsymbol{x}$ 出发作主成份分析所得到的结果是不同的.

克服上述困难的一个办法是将 x 归 1 化，即令 $\boldsymbol{x}^*=\boldsymbol{x}\boldsymbol{D}$，$\boldsymbol{D}=\mathrm{diag}\left(\frac{1}{\sigma_1},\cdots,\frac{1}{\sigma_p}\right)$，其中 σ_i^2 为 x_i 的方差. 这样，$\boldsymbol{x}^*$ 的各分量有方差 1，而各分量之间的协方差就是相关系数. 从 $\boldsymbol{x}^*$ 出发作主成份分析相当于在定义 18.6.1 中用 $\boldsymbol{x}$ 的相关阵代替方差阵 $\boldsymbol{\Sigma}$，而在主成份的表达式中用 $\boldsymbol{x}^*$ 去代替 $\boldsymbol{x}$. 这样做的好处是不致于因某些重要的变量的观测数据的绝对值过小而被“淹没”掉；其缺点是有可能“放大”观测误差，从而加重某些不重要变量的份量.

18.6.2 主成份的应用

首先我们给出一个主成份分析的实例.

例 18.6.5 贝塞尔(Berce)和维尔博(Wilbaut)收集了在11年中5个气象变量的观测数据如表18.1.

表 18.1

年度	x_1	x_2	x_3	x_4	x_5
1920—1921	87.9	19.6	1.0	1661	28.37
1921—1922	89.9	15.2	90.1	968	23.77
1922—1923	153.0	19.7	56.6	1353	26.04
1923—1924	132.1	17.0	91.0	1293	25.74
1924—1925	88.8	18.3	93.7	1153	26.68
1925—1926	220.9	17.8	106.9	1286	24.29
1926—1927	117.7	17.8	65.5	1104	28.00
1927—1928	109.0	18.3	41.8	1574	28.37
1928—1929	156.1	17.8	57.4	1222	24.96
1929—1930	181.5	16.8	140.6	902	21.66
1930—1931	181.4	17.0	74.3	1150	24.37

其中5个变量的含义是:

x_1=11至12月份的降雨量(mm),

x_2=7月份的平均气温(C),

x_3=7月份的降雨量(mm),

x_4=7月份的辐射度(mm 酒精),

x_5=平均产量(公担/英亩).

由以上数据算得方差和相关矩阵如下

$$
\begin{bmatrix}
1794 & -4.473 & 726.9 & -2218 & -52.01 \\
-0.0087 & 1.488 & -26.62 & 197.5 & 1.577 \\
0.491 & -0.624 & 1224 & -6203 & -56.44 \\
-0.239 & 0.738 & -0.808 & 48104 & 328.9 \\
-0.607 & 0.640 & -0.798 & 0.742 & 4.087
\end{bmatrix}
\tag{18.108}
$$

在此矩阵中，主对角线上是方差；主对角线以上是方差；主对角线以下是相关系数.

对前 4 个变量(x_1-x_4)作主成份分析.由方差矩阵出发算得 4 个特征根依次为

$$\lambda_1=49023,\lambda_2=1817,\lambda_3=283 \text{ 和 } \lambda_4=0.6. \tag{18.109}$$

相应的特征向量为

$$\begin{array}{l}(-0.049,\quad 0.004,-0.129,\quad 0.990),\\(-0.954,\quad 0.003,\quad 0.288,\quad 0.084),\\(-0.296,-0.008,\quad 0.949,\quad 0.109),\\(-0.005,\quad 0.999,\quad 0.008,-0.003).\end{array} \tag{18.110}$$

反之，由相关矩阵出发算得 4 个特征根依次为

$$\lambda_1=2.60,\lambda_2=0.96,\lambda_3=0.28,\lambda_4=0.16. \tag{18.111}$$

而相应的特征向量为

$$\begin{array}{l}(\quad 0.291,-0.506,\quad 0.577,-0.571),\\(\quad 0.871,\quad 0.425,\quad 0.136,\quad 0.205),\\(-0.332,\quad 0.742,\quad 0.418,-0.405),\\(-0.214,-0.111,\quad 0.688,\quad 0.685).\end{array} \tag{18.112}$$

由以上计算结果可以看出，第 2 个变量(x_2)在由方差阵计算的主成份中几乎全部集中在第 4 个主成份中(系数几乎为 1)，而在前三个主成份中所占比重几乎为零(系数分别为 0.004，0.003 和 -0.008).相反，在由相关阵计算的主成份中，标准化的变量 x_2^* 在前三个主成份中都占有相当的比重.从问题的实际背景来看，显然后一分析结果更符合实际.

不论是从方差阵出发还是从相关阵出发，主成份分析都是通过对方差阵或相关阵进行谱分解进行的.下面统一用 $\boldsymbol{\Sigma}$ 记方差阵或相关阵，用 $\boldsymbol{x}$ 记原变量或标准化变量.我们来考虑 $\boldsymbol{x}$ 与主成份 $\boldsymbol{y}$ 之间的相关性.记

$$\boldsymbol{\Sigma}=\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\Gamma},\boldsymbol{y}=\boldsymbol{x}\boldsymbol{\Gamma}, \tag{18.113}$$

其中 Γ 为正交阵，$\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_p)$. 我们有

$$\text{cov}(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}^{\mathrm{T}})=\Sigma\Gamma=\Gamma\Lambda. \tag{18.114}$$

因此变量 x_i 与主成份 y_i 之间的相关系数为

$$\rho_{ij}=\frac{\text{cov}(x_i,y_j)}{(V(x_i)V(y_j))^{\frac{1}{2}}}=\gamma_{ij}\sqrt{\frac{\lambda_j}{\sigma_{ii}}}, \tag{18.115}$$

其中 γ_{ij} 为 Γ 的第 i 行第 j 列元素，σ_{ii} 为 x_i 的方差. 特别，当 $\boldsymbol{x}$ 为标准化变量时有

$$\rho_{ij}=\gamma_{ij}\sqrt{\lambda_j}. \tag{18.116}$$

容易验证

$$\sum_{j=1}^{p}\rho_{ij}^2=1,\quad i=1,\cdots,p. \tag{18.117}$$

定义 18.6.6 称 ρ_{ij}^2 为用主成分 y_j 解释的变量 x_i 的**变差比例**.

对例 18.6.5 中前 4 个变量 $x_1,\cdots,x_4$ 与由之产生的主成份 $y_1,\cdots,y_4$（用相关阵，见(18.112)），我们用公式(18.116)可得到它们之间的相关系数阵如下

	y_1	y_2	y_3	y_4
x_1	0.468	0.862	−0.177	−0.081
x_2	−0.815	0.420	0.397	−0.042
x_3	0.930	0.135	0.223	0.260
x_4	−0.919	0.202	−0.216	0.259

(18.118)

这样，对变量 x_1，第 1 主成份解释它的变差 $(0.468)^2\times100\%=21.9\%$，第 2 主成份解释它的变差 $(0.862)^2\times100\%=74.3\%$，…，等等.

在被研究变量的个数(p)相当大时，研究数据的结构，发现其特征是较困难的. 利用主成份分析可以降低数据的维数，具体的作法是选取适当的 $k(<p)$，用前 k 个主成份的观测数据来代替原来

的数据阵. 如何选定 k 呢?这就要看前 k 个主成份能在多大程度上来解释原来的 p 个变量. 用定义 18.6.6 知 x_i 的方差 σ_{ii} 中被第 j 个主成份解释的部分为 $\sigma_{ii}\rho_{ij}^2$. 我们容易证明

$$\sum_{i=1}^{p}\sigma_{ii}\rho_{ij}^2 = \lambda_j, \tag{18.119}$$

$$\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma_{ii}\rho_{ij}^2 = \sum_{j=1}^{p}\lambda_j = \sum_{i=1}^{p}\sigma_{ii} = \mathrm{tr}(\boldsymbol{\Sigma}). \tag{18.120}$$

因此,我们如用 $\mathrm{tr}(\boldsymbol{\Sigma})$来表示原来 p 个变量的"总变差"的话,则

$$\frac{\lambda_j(\text{第 } j \text{ 个主成份的方差})}{\sum_{j=1}^{p}\lambda_j(\text{总变差})} \tag{18.121}$$

就表示第 j 个主成份所解释的变差在总变差中的比例. 当我们选前 k 个主成份来代替原来的 p 个变量时,则这 k 个主成份所解释的变差在总变差中的比例为

$$\frac{\sum_{j=1}^{k}\lambda_j}{\sum_{j=1}^{p}\lambda_j}. \tag{18.122}$$

因此可以根据问题的需要事先给出一个阀值 λ_T(例如 $\lambda_T=90\%$或95%等等),当

$$\left(\sum_{j=1}^{k}\lambda_j\right)\Big/\left(\sum_{j=1}^{p}\lambda_j\right) \geqslant \lambda_T, \text{而} \tag{18.123}$$

$$\left(\sum_{j=1}^{k-1}\lambda_j\right)\Big/\left(\sum_{j=1}^{p}\lambda_j\right) < \lambda_T \text{ 时}$$

就取前 k 个主成份.

另外一种选择主成份的方法是:

丢弃那些相应的特征根小于$\left(\sum_{j=1}^{k}\lambda_j\right)\Big/p$ 的那些主成份. 特别

当 Σ 为相关阵时，丢弃那些相应的特征值小于 1 的主成份.

根据实践经验，用前一种方法往往会保留过多的主成份而用后一种方法往往丢掉过多的主成份. 因此，可以根据实际问题的需要从上述两种方法中折中，选取一个适当的 k.

例 18.6.7 在例 18.6.5 中考虑用相关阵计算的主成份. 用(18.123)的方法，取 $\lambda_t=90\%$，则 $k_1=3$，即取前 3 个主成份，而用后一种方法得到 $k_2=1$，即只取第 1 个主成份. 折中一下，取 $k=2$ 是适当的.

在实际问题中，主成份（特别是前几个主成份）往往有一定的实际含义. 如何解释主成份的含义要根据具体问题的客观背景去分析，不是一两句话能说清楚的. 下面介绍一种主成份在解释数据的尺度和形状方面的应用.

假定第一主成份中各变量的系数都是正的，则由正交性一定可以推出其它主成份的系数是有正有负的，于是第一主成份便可解释为**尺度因子**(scale factor 或 size factor). 而其它被选中的主成份可解释为不同的**形状因子**(shape factor(s)). 这种情形的出现可有下述定理做保证.

定理 18.6.8 若方差矩阵的所有元素均为非负数，则第 1 主成份的系数均为正值.

例 18.6.9 对 100 名学生在 5 门课的考试中的成绩进行统计，这 5 门课程的考试有闭卷考试，有开卷考试，具体内容如下所列：

x_1＝力学(闭)，

x_2＝向量(闭)，

x_3＝代数(开)，

x_4＝分析(开)，

x_5＝统计(开).

样本协方差阵为

$$
\begin{bmatrix}
302.3 & 125.8 & 100.4 & 105.1 & 116.1 \\
 & 170.9 & 84.2 & 93.6 & 97.9 \\
 & & 111.6 & 110.8 & 120.5 \\
 & & & 217.9 & 153.8 \\
 & & & & 294.4
\end{bmatrix}
\tag{18.124}
$$

算得样本方差阵的特征根分别为

$$
\lambda_1 = 679.2, \lambda_2 = 199.8, \lambda_3 = 102.6, \lambda_4 = 83.7, \lambda_5 = 31.8. \tag{18.125}
$$

要选择 k 个主成份，根据第一种方法（见 18.6.23），若定 $\lambda_T = 0.9$，则 $k_1 = 4$；按第二种方法得 $k_2 = 1$，因此经过折中可考虑 $k = 2$ 或 3. 在这里为使问题简化定 $k = 2$，于是前两个主成份为

$$
\begin{aligned}
y_1 &= 0.51x_1 + 0.37x_2 + 0.35x_3 + 0.45x_4 + 0.53x_5, \\
y_2 &= 0.75x_1 + 0.21x_2 - 0.08x_3 - 0.30x_4 - 0.55x_5.
\end{aligned}
\tag{18.126}
$$

这里第 1 主成份的系数均为正，因此可作尺度因子. 将第 1 主成份的各项系数乘上一个相同的因数，使之和为 1，得到

$$
y'_1 = 0.23x_1 + 0.17x_2 + 0.16x_3 + 0.20x_4 + 0.24x_5. \tag{18.127}
$$

y'_1 可作为学生的平均成绩，用来刻划学生的学习总效果. 对第二主成份注意 x_1 和 x_2 的系数为正而这两门均为闭卷考试，但 x_3，x_4 和 x_5 的系数为负而这三门均为开卷考试，因此 y_2 作为闭卷考成绩与开卷考成绩之差，可用来检查学生应付开卷考或闭卷考能力的度量. 以某个学生来说，y_2 的值大表示他（她）应付闭卷考的能力比应付开卷考的能力大；y_2 小则情形相反. 下面列出 5 个学生的成绩及 100 个学生的总平均成绩如表 18.2.

表 18.2

学生	x_1	x_2	x_3	x_4	x_5	y_1	y_2
1	77	82	67	67	81	75.2	5.0
2	63	78	80	70	81	74.0	−8.3
3	75	73	71	66	81	73.7	1.6
4	55	72	63	70	68	65.3	−7.1
5	31	55	60	57	73	55.0	−27.3
$\bar{x}/\bar{y}$	39.0	50.6	50.6	46.7	42.3	45.2	1.4

从表 18.2 中可以一目了然地看出这 5 个学生的总平均成绩及应付开闭卷考试的能力(与总平均成绩相比较).

19 多元分析(Ⅱ)

19.1 典则相关分析

19.1.1 典则相关变量

典则相关分析(canonical correlation analysis)是用于分析两个随机向量之间相关性程度的一种统计方法. 设 $\boldsymbol{x}=(x_1,\cdots,x_p)^{\mathrm{T}}$ 和 $\boldsymbol{y}=(y_1,\cdots,y_q)^{\mathrm{T}}$ 为两个随机向量. 如何分析 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的相关性呢? 当 $p=q=1$ 时,显然可以用 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的相关系数 $\rho(\boldsymbol{x},\boldsymbol{y})$ 来描述,而当 p 和 q 之中有一个大于 1 时,就不能直接计算 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的相关系数,但 $\boldsymbol{x}$ 的线性函数与 $\boldsymbol{y}$ 的线性函数之间的相关系数是可以计算的,因此可以考虑分别选择 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的适当的线性函数,用它们之间的相关性程度来刻划 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的相关程度.

设 $\boldsymbol{a}$ 为任一 p 维常向量,$\boldsymbol{b}$ 为任一 q 维常向量.

定义 19.1.1 设 p 维常向量 $\boldsymbol{a}_1$ 和 q 维常向量 $\boldsymbol{b}_1$ 满足:

$$|\rho(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{y})| = \max_{\substack{\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})=1 \\ \sigma^2(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y})=1}} |\rho(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y})|.$$

则称 $\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x}$ 和 $\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{y}$ 为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的第一组**典则相关变量**,$|\rho(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{y})|$ 称为相应的**典则相关系数**.

由定义,第一组典则相关变量之间的相关系数在所有关于 x 和 y 的线性函数之中达到最大.

但仅用一组典则相关变量来刻划随机向量之间的关系还是不够的,进一步可以考虑在所有与 $\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x}$ 不相关的 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ 和所有与 $\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{y}$ 不相关的 $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$ 中挑选 $\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x}$ 和 $\boldsymbol{b}_2^{\mathrm{T}}\boldsymbol{y}$,,它们之间的相关系数达到最大,等等.

定义 19.1.2 对 $k\geqslant 2$，设第 $k-1$ 组典则相关变量已经定义，设 $\boldsymbol{a}_k$ 和 $\boldsymbol{b}_k$ 分别为 p 维和 q 维常向量，满足：

$$|\rho(\boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{y})| = \max_k|\rho(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y})|,$$

其中 $\max\limits_k$ 的含义是在 $\sigma^2(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})=\sigma^2(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}))=1$ 及 $\rho(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x},\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{x})=\rho(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y},\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{y})=0,j=1,\cdots,k-1$ 的条件下求极大，则称 $\boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{x}$ 和 $\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{y}$ 为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的**第 k 组典则相关变量**. $|\rho(\boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{y})|$ 称为**第 k 个典则相关系数**.

由以上定义知，典则相关变量的组数 k 不超过 $\min(p,q)$，且 $|\rho(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{y})|\geqslant\rho|(\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_2^{\mathrm{T}}\boldsymbol{y})|\geqslant\cdots\geqslant|\rho(\boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{y})|$. 假定事先指定一个阈值 ρ_0，则可以根据 $|\rho(\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{y})|$ 的大小将 k 组典则相关变量分成两类，第一类满足 $|\rho(\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{y})|\geqslant\rho_0$，代表 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 相关程度较大的那部分，第二类满足 $|\rho(\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{x},\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{y})|<\rho_0$，代表 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 相关程度较小的那部分.

19.1.2 典则相关变量的计算

设随机向量 $(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}_1^{\mathrm{T}})^{\mathrm{T}}$ 的方差阵为

$$\boldsymbol{V}=\boldsymbol{V}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}=\begin{pmatrix}V(x)\mathrm{cov}(xy)\\ \mathrm{cov}(y,x)V(y)\end{pmatrix}=\begin{pmatrix}\boldsymbol{V}_{xx}\,\boldsymbol{V}_{xy}\\ \boldsymbol{V}_{yx}\,\boldsymbol{V}_{yy}\end{pmatrix}. \qquad (19.1)$$

定理 19.1.3 设 $p\geqslant q$，且 $\boldsymbol{V}_{yy}$ 和 $\boldsymbol{V}_{xx}$ 均为非奇异阵，对称正定阵 $\boldsymbol{V}_{yy}^{-\frac{1}{2}}\boldsymbol{V}_{yx}\boldsymbol{V}_{xx}^{-1}\boldsymbol{V}_{xy}\boldsymbol{V}_{yy}^{-\frac{1}{2}}$ 有特征根 $\eta_1\geqslant\eta_2\geqslant\cdots\geqslant\eta_q>0$，$\boldsymbol{u}_i$ 为相应于 η_i 的单位化特征向量，$i=1,\cdots,q$，令

$$\boldsymbol{a}_i=\eta_i^{-\frac{1}{2}}\boldsymbol{V}_{xx}^{-1}\boldsymbol{V}_{xy}\boldsymbol{b}_i,\quad \boldsymbol{b}_i=\boldsymbol{V}_{yy}^{-\frac{1}{2}}\boldsymbol{u}_i,\quad i=1,2,\cdots,q.$$

则 $\boldsymbol{a}_i'\boldsymbol{x},\boldsymbol{b}_i'\boldsymbol{y}$ 为第 i 组典则相关变量，$\eta_i^{\frac{1}{2}}$ 为相应的典则相关系数.

在上述定理中，当 $p<q$ 时，可反过来求 $\boldsymbol{V}_{xx}^{-\frac{1}{2}}\boldsymbol{V}_{xy}\boldsymbol{V}_{yy}^{-1}\boldsymbol{V}_{yx}\boldsymbol{V}_{xx}^{-\frac{1}{2}}$ 的特征根 λ_i 和特征向量 $\boldsymbol{v}_i$，令 $\boldsymbol{a}_i=\boldsymbol{V}_{xx}^{-\frac{1}{2}}\boldsymbol{v}_i,\boldsymbol{b}_i,=\lambda_i^{-\frac{1}{2}}\boldsymbol{V}_{yy}^{-1}\boldsymbol{V}_{yx}\boldsymbol{a}_i,i=1,2,\cdots,p$.

在实际问题中，一般不知道方差阵 $\boldsymbol{V}$，而是有 $\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}$ 的 n 个样本 $\begin{pmatrix}\boldsymbol{x}_{(1)}\\ \boldsymbol{y}_{(1)}\end{pmatrix},\begin{pmatrix}\boldsymbol{x}_{(2)}\\ \boldsymbol{y}_{(2)}\end{pmatrix},\cdots,\begin{pmatrix}\boldsymbol{x}_{(n)}\\ \boldsymbol{y}_{(n)}\end{pmatrix}$，计算 $\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}$ 的样本方差阵

$$\boldsymbol{S}=\begin{pmatrix}\boldsymbol{S}_{xx} & \boldsymbol{S}_{xy}\\ \boldsymbol{S}_{yx} & \boldsymbol{S}_{yy}\end{pmatrix}.$$

将定理 19.1.3 中的 $\boldsymbol{V}_{xx},\boldsymbol{V}_{xy},\boldsymbol{V}_{yx},\boldsymbol{V}_{yy}$ 分别换成 $\boldsymbol{S}_{xx},\boldsymbol{S}_{xy},\boldsymbol{S}_{yx},\boldsymbol{S}_{yy}$，则可得到样本形式的典则相关变量 $\hat{\boldsymbol{a}}_i^{\mathrm{T}}\boldsymbol{x},\hat{\boldsymbol{b}}_i^{\mathrm{T}}\boldsymbol{y}$ 和典则相关系数 $\hat{\eta}_i^{\frac{1}{2}},i=1,\cdots,q$.

大样本理论证明：当 $n\to\infty$ 时有 $\hat{\boldsymbol{a}}_i\xrightarrow{P}\boldsymbol{a}_i,\hat{\boldsymbol{b}}_i\xrightarrow{P}\boldsymbol{b}_i$ 及 $\hat{\eta}_i\xrightarrow{P}\eta_i$，$i=1,\cdots,q$.

19.2 因子分析

因子分析(factor analysis)是一种用来分析隐藏在表象背后的因子作用的一类统计模型和方法，它起源于**心理度量学**(phsycholometrics)，在方法上与主成份分析有密切联系.

19.2.1 正交因子模型

设 $\boldsymbol{x}$ 为一个 p 维可观测随机向量，假定 $\boldsymbol{x}$ 受到 m 个不可观测的随机因子的控制，称这 m 个影响 $\boldsymbol{x}$ 的因子为公共因子，用 m 维随机向量 $\boldsymbol{f}$ 来表示，又假定 $\boldsymbol{f}$ 对 $\boldsymbol{x}$ 的影响是线性的，则 $\boldsymbol{x}$ 与 $\boldsymbol{f}$ 之间的关系可用下述模型来表述：

模型 19.2.1 $\quad \boldsymbol{x}=\boldsymbol{\mu}+\boldsymbol{L}\boldsymbol{f}+\boldsymbol{\varepsilon}$

其中 $\boldsymbol{\mu}$ 为 p 维常向量，表示 $\boldsymbol{x}$ 的均值；$\boldsymbol{L}$ 为 $p\times m$ 维常数阵，$\boldsymbol{L}$ 的第 i 行表示公共因子 $\boldsymbol{f}$ 对 $\boldsymbol{x}$ 第 i 个分量 x_i 的影响系数；$\boldsymbol{\varepsilon}$ 为 p 维随机变量，代表 $\boldsymbol{x}$ 中与 $\boldsymbol{f}$ 无关的那一部分，称为特殊因子.

在模型 19.2.1 中，$\boldsymbol{f}$ 和 $\boldsymbol{\varepsilon}$ 都是不可观测的，假定它们满足下

述条件：

（1）$\boldsymbol{f}$ 与 $\boldsymbol{\varepsilon}$ 相互独立. （19.2）

（2）$E(\boldsymbol{f})=0$，$V(\boldsymbol{f})=\boldsymbol{I}_m$.

（3）$E(\boldsymbol{\varepsilon})=0$，$V(\boldsymbol{\varepsilon})=\boldsymbol{\Psi}$. $\boldsymbol{\Psi}$ 为对角阵.

在上述条件中，要求 $V(\boldsymbol{f})=\boldsymbol{I}_m$，即要求各因子之间不相关，这样的模型便称为**正交因子模型**. 在正交因子模型中，公共因子 $\boldsymbol{f}$ 对 $\boldsymbol{x}$ 的各分量都起作用，而特殊因子 $\boldsymbol{\varepsilon}$ 的第 i 个分量 ε_i 只对 $\boldsymbol{x}$ 的第 i 个分量 x_i 起作用. 模型 19.2.1 中矩阵 $\boldsymbol{L}$ 称为**载荷阵**（loading matrix），是有待确定的.

在模型 19.2.1 中，假定将 $\boldsymbol{f}=\boldsymbol{Th}$ 代入，其中 $\boldsymbol{T}$ 为正交矩阵，$\boldsymbol{h}$ 为另一个 m 维随机向量，则有 $\boldsymbol{x}=\boldsymbol{Hh}+\boldsymbol{\varepsilon}$，其中 $\boldsymbol{H}=\boldsymbol{LT}$. 容易验证，在（19.2）中，用 $\boldsymbol{h}$ 代替 $\boldsymbol{f}$ 后，各项条件仍然成立. 因此，正交因子模型是不确定的. 这种不确定性虽然在一定程度上造成了求解的困难，但另一方面也提供了更多的可选择性.

引理 19.2.2 在模型 19.2.1 中，记 $\boldsymbol{L}=(l_{ij})$，$\boldsymbol{\Psi}=\mathrm{diag}(\psi_1,\cdots,\psi_p)$，由条件（19.2）可得

（1）$\boldsymbol{V}(x)=\boldsymbol{LL}'+\boldsymbol{\Psi}$，即

$$\sigma^2(x_i)=\sum_{j=1}^{m}l_{ij}^2+\Psi_i,\quad i=1,\cdots,p.$$

（2）$\mathrm{cov}(\boldsymbol{x},\boldsymbol{f})=\boldsymbol{L}$，或

$$\mathrm{cov}(x_i,f_j)=l_{ij},\quad i=1,\cdots,p,\ j=1,\cdots,m.$$

在 x_i 的方差中，通常称由公共因子贡献的部分 $\left(\sum_{j=1}^{m}l_{ij}^2\right)$ 为**公共性**（communality）；而由特殊因子贡献的部分（ψ_i）称为**唯一性**（uniqueness）. l_{ij} 在当 $\sigma^2(x_i)=1$ 时表示 x_i 与 f_j 的相关程度，称为 x_i 在 f_j 上的**载荷**（loading）. 而

$$g_j^2=\sum_{i=1}^{p}l_{ij}^2 \tag{19.3}$$

刻划了 f_j 对所有 p 个变量影响的程度，称为 f_j 对 $\boldsymbol{x}$ 的**贡献**

(contribution).

上述关系在分析因子结构时有重要作用.

19.2.2 估计方法

在因子模型19.2.1中,关键是要估计载荷阵$\boldsymbol{L}$和特殊因子方差阵$\boldsymbol{\Psi}$.设$\boldsymbol{x}$有n个观测样本,并据此已计算出样本方差阵$\boldsymbol{S}$和相关阵$\boldsymbol{R}$.通常从样本相关阵$\boldsymbol{R}$出发进行计算,这相当于已将$\boldsymbol{x}$各分量的方差都取成1了.从模型上看,这样做相当于将19.2.1中$\boldsymbol{L}$改为$\boldsymbol{D}^{-1}\boldsymbol{L}$,其中$\boldsymbol{D}=\mathrm{diag}(\sigma(x_1),\cdots,\sigma(x_p))$,而将$\boldsymbol{\varepsilon}$改为$\tilde{\boldsymbol{\varepsilon}}=\boldsymbol{D}^{-1}\boldsymbol{\varepsilon}$,而条件(19.2)仍然成立(将$\boldsymbol{\varepsilon}$改为$\tilde{\boldsymbol{\varepsilon}}$).下面介绍两种估计$\boldsymbol{L}$和$\boldsymbol{\Psi}$的方法.

方法19.2.3 主成份法.由引理19.2.2中的(1)式,设$\hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_p$是样本相关阵$\boldsymbol{R}$的特征根,$\hat{\boldsymbol{l}}_1,\cdots,\hat{\boldsymbol{l}}_p$为相应的单位特征向量.对给定的公因子数$m(<p)$,对因子载荷阵$L$的估计为

$$\hat{\boldsymbol{L}}=\left(\sqrt{\hat{\lambda}_1}\hat{\boldsymbol{l}}_1\vdots\sqrt{\hat{\lambda}_2}\hat{\boldsymbol{l}}_2\vdots\cdots\vdots\sqrt{\hat{\lambda}_m}\hat{\boldsymbol{l}}_m\right),\tag{19.4}$$

特殊方差的估计为

$$\hat{\boldsymbol{\Psi}}=\mathrm{diag}(\hat{\psi}_1,\cdots,\hat{\psi}_p),\tag{19.5}$$

其中$\hat{\psi}_i=r_{ii}-\sum\limits_{j=1}^{m}\hat{l}_{ij}^2,\quad i=1,\cdots,p.$

r_{ii}为$\boldsymbol{R}$的第i个对角元素.

在主成份方法中,相当于用前m个主成份去表示潜在的m个因子.这相当于给主成份以一定的物理解释.

在实际问题中,因子的个数m通常是有待确定的.确定m的方法通常有以下几种:

取m,使得$\boldsymbol{R}-\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$接近对角阵,即$\boldsymbol{R}-\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$的非对角元接近零. (19.6)

取 m，使得 $\sum_{j=1}^{m}\hat{\lambda}_j\Big/\sum_{j=1}^{p}\hat{\lambda}_j \geqslant 85\%$. (19.7)

取 m，使得 $\hat{\lambda}_j\geqslant 1, j=1,\cdots,m$; $\hat{\lambda}_j<1, j=m+1,\cdots,p$. (19.8)

在实际问题中可用不同的方法确定 m，然后从中取一个与实际问题背景最接近的解.

方法 19.2.4 极大似然法. 假定公共因子 $\boldsymbol{f}$ 和特殊因子 $\boldsymbol{\epsilon}$ 均为正态分布，则 $\boldsymbol{x}$ 的分布为

$$N_p(\mu_1,\boldsymbol{\Sigma})=N_p(\mu_1,\boldsymbol{LL}^{\mathrm{T}}+\boldsymbol{\Psi}). \tag{19.9}$$

对 $\boldsymbol{x}$ 的独立同分布样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$，可求得其对数似然函数为

$$\log\boldsymbol{L}(\mu,\boldsymbol{\Sigma}) = -\frac{np}{2}\log(2\pi)-\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}+n(\bar{x}-\mu)(\bar{x}-\mu)^{\mathrm{T}}], \tag{19.10}$$

其中 $\bar{x}$ 为样本均值，$\boldsymbol{S}$ 为样本方差阵. 由于 $\boldsymbol{\Sigma}=\boldsymbol{LL}^{\mathrm{T}}+\boldsymbol{\Psi}$，$\log\boldsymbol{L}(\mu,\boldsymbol{\Sigma})$ 依赖于 $\boldsymbol{L}$ 和 $\boldsymbol{\Psi}$. 为计算方便起见，对 $\boldsymbol{L},\boldsymbol{\Psi}$ 加如下约束

$$\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\boldsymbol{L}=\boldsymbol{\Delta}, \tag{19.11}$$

其中 $\boldsymbol{\Delta}$ 为对角阵，在此约束下，求

$$\min\log\boldsymbol{L}(\mu,\boldsymbol{\Sigma}) \tag{19.12}$$

的解. 上述求解过程可通过非线性优化的方法寻找数值解.

在上一小节中提到，如果 $\boldsymbol{L}$ 为一个载荷阵，则对任一正交矩阵 $\boldsymbol{T}(\boldsymbol{TT}^{\mathrm{T}}=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}=\boldsymbol{I}_m)$，$\boldsymbol{L}^*=\boldsymbol{LT}$ 仍是一个因子载荷阵，相当将原来的公共因子 $\boldsymbol{f}$ 作一个正交变换后变到 $\boldsymbol{h}=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{f}$. 这一性质给我们提供了寻找“理想”因子结构的思路. 所谓“理想”的因子结构就是说，我们希望看到：每个变量的载荷都尽可能集中在某个因子上，而在其他因子上的载荷尽可能小. 根据上述思路，凯瑟(Kaiser)于 1958 年提出一个寻找理想因子结构的准则，称为**方差最大准则**

(varimax criterion);记 $\hat{h}_i = \sum_{j=1}^{m} \hat{l}_{ij}^2, i = 1, \cdots, p$. 记 $\hat{l}_{ij}^*$ 为经过旋转后的载荷,用

$$V_j = \frac{1}{p}\left[\sum_{i=1}^{p} (\hat{l}_{ij}^* / h_i)^4 - \left(\frac{1}{p} \sum_{i=1}^{p} \frac{(\hat{l}_{ij}^*)^2}{h_i^2} \right)^2 \right] \quad (19.13)$$

来度量第 j 个因子在 p 个变量上载荷的离散度. 理论上说 V_j 应越大越好,$j=1,\cdots,m$. 从总体上来讲,要求

$$V = \sum_{j=1}^{m} V_j \quad (19.14)$$

达到最大. 具体的算法有各种数值方法可采用.

19.3 聚类分析

聚类分析(cluster analysis)是一类名目繁多的将数据所对应的研究对象(客体)进行分类的统计方法. 这一类方法的共同特点是:事先不知道类别的个数与结构;据以进行分析的数据是客体之间的**相似性**(similarity)或**相异性**(dissimilarity)数据. 将这些相似(相异)性数据看成是客体之间"距离"远近的一种度量,将距离近的客体归入同一类,不同类之间的客体距离较远. 这就是聚类分析方法的共同思路.

19.3.1 相似性与距离

相似性数据是用于度量客体之间相似程度的数据. 设有两个标号分别为 i 和 j 的客体,用 c_{ij} 表示客体 i 和 j 之间的相似性数据,则 c_{ij} 愈大反映客体之间愈相似,从聚类的观点来看,意味着客体 i 与 j 之间"距离"较小;反之,c_{ij} 小,i 与 j 之间的相似程度小,因而认为 i 与 j 间"距离"大.

对相似性数据的基本要求是:

$$c_{ii} > c_{ij}, \quad i \neq j, \quad (19.15)$$

理由是显而易见的. 为便于使用，通常还要求相似性数据满足：

对称性　　$c_{ij}=c_{ji}$，　　(19.16)

及

$$|c_{ij}|\leqslant 1. \tag{19.17}$$

相似性数据的来源有两种，一种来源是直接得到的. 例如在心理学试验中，常令若干个试验者为两两成对的研究客体的相似程度打分，将这些得分进行加权平均，便得到一组相似性数据.

相似性数据的另一个来源是由通常的一组 n 个样品 p 个变量的观测数据通过一定的计算公式得到. 这时原始数据是一个 $n\times p$ 矩阵 $\boldsymbol{X}$，每行对应一个样品，每列对应一个变量. 这时客体(objects)，即聚类的对象，可以是样品(这时有 n 个客体)，也可以是变量(这时有 p 个客体). 如果是以样品为客体，则由 $\boldsymbol{X}$ 产生出一个 $n\times n$ 的对称阵 $\boldsymbol{C}$，其中元素 c_{ij}表示样品 i 与样品 j 之间的相似性，$\boldsymbol{C}$ 称为相似阵. 如果以变量为客体，则由 $\boldsymbol{X}$ 产生出一个 $p\times p$ 的相似阵. 相似性数据的计算方法根据以样品为客体或以变量为客体会有所不同. 除此以外，测量原始数据的度量尺度不同也影响相似性的计算公式. 由于以上原因，相似性数据的计算公式名目繁多，本书不能详细介绍，请读者参阅[18].

相异性数据是客体之间相异程度的度量. 从聚类的观点看，相异性愈大，反映客体之间距离愈远；反之，客体间距离近. 用 δ_{ij}记客体 i 与客体 j 之间的相异性. 则基本的要求是：

$$\delta_{ij}\geqslant 0,\ \delta_{ij}=0\Leftrightarrow i=j. \tag{19.18}$$

$$\delta_{ij}=\delta_{ji}. \tag{19.19}$$

以上两条性质与数学意义上距离的性质相同，但相异性不必满足距离的第三条性质：

$$\delta_{ij}\leqslant\delta_{ik}+\delta_{jk}. \tag{19.20}$$

在聚类分析中，不论相异性是否满足(19.20)式，都广义地将相异性看成是距离. 另外，相似性数据也可以通过适当的变换转化

为相异性(距离).例如可考虑用下述变换:

$$\delta_{ij} = 1 - c_{ij}, \tag{19.21}$$

或

$$\delta_{ij} = [c_{ii} - 2c_{ij} + c_{jj}]^{\frac{1}{2}}. \tag{19.22}$$

等.

关于相似性、相异性和距离的各种定义方法均可参考[18].以下将统一使用距离这个概念来概括相似性或相异性,并在此基础上讨论聚类的方法.

19.3.2 层次聚类法

层次聚类(hierarchical clustering)是一类聚类方法.这类方法的基本特征是:先将 m 个客体各自看成一类,是初始分类,记作 C_0.然后根据类与类之间的距离来合并,将 C_0 中距离最近的类合并成一类,得到一组新的类,记为 C_1.在 C_1 这层分类的基础上,再重复以上 $C_0 \to C_1$ 的过程,得到一组新的类,记为 C_2,以上聚类过程可以不断重复进行,得到一个聚类层的序列 $C_0, C_1, C_2, \cdots, C_k$.其中 C_k 中只有一类,即包含所有 m 个客体的类.以上聚类过程可用**聚类树**(clustering tree)来图示(图 19.1).

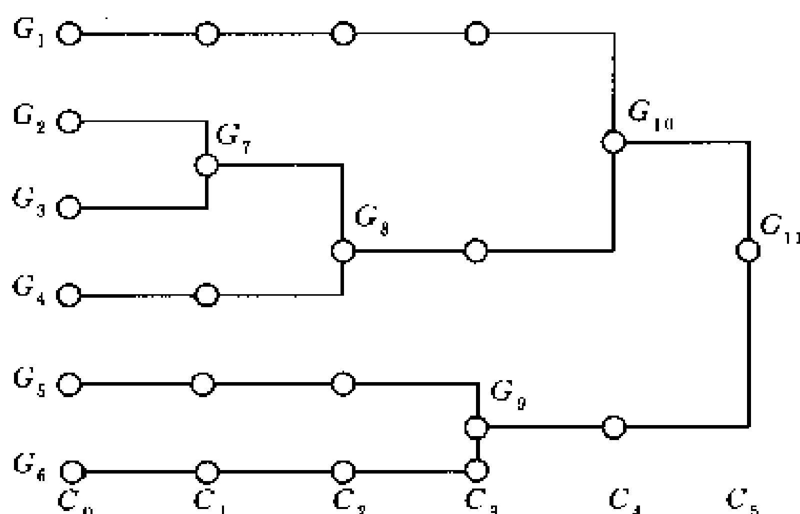


图 19.1 聚类树

图中显示了 6 个客体的分层聚类过程，得到 6 个聚类层 C_0，C_1，…，C_5，每一层是将上一层的两个类合并为一个新类得到的. 当然，从原则上说，在层次聚类中，每一个新类可以由两个以上的旧类合并而成；且每一层可以得到不止一个新类. 但从计算简便，易于编制计算机程序的角度来看，每次得一个新类，每个新类由两个旧类合并而成的方法较好.

在层次聚类过程中，关键是要定义一个合适的类间距离. 通常类间距离是点间距离的函数. 类间距离的不同定义决定了各种层次聚类的不同特点. 因此常用类间距离的名称命名层次聚类的方法. 下面介绍几种常用的层次聚类法.

方法 19.3.1　最短距离法(shortest distance method)　用 d_{ij}表示客体 i 和客体 j 之间的距离，G_1，G_2，…表示类，用 D_{pq}表示类 G_p 与类 G_q 之间的距离，则

$$D_{pq} = \min_{i \in G_p, j \in G_q} d_{ij},$$

即类间距离为两类间最近点的距离. 设 $G_r = \{G_p, G_q\}$，则 G_r 与另一个类 G_k 之间的距离有递推公式

$$D_{rk} = \min\{D_{pk}, D_{qk}\}.$$

方法 19.3.2　最长距离法(longest distance method)　与最短距离法相反，最长距离法定义类间距离为两类之间最远点的距离，即

$$D_{pq} = \max_{i \in G_p, j \in G_q} d_{ij}.$$

类间距离的递推公式为(当 $G_r = \{G_p, G_q\}$)

$$D_{rk} = \max\{D_{pk}, D_{qk}\}.$$

方法 19.3.3　中间距离法(median method)　中间距离法是对最短距离法和最长距离法的一个折衷，即令类间距离的递推公式为(当 $G_r = \{G_p, G_q\}$)：

$$D_{kr}^2 = \frac{1}{2} D_{kp}^2 + \frac{1}{2} D_{kq}^2 - \frac{1}{4} D_{pq}^2.$$

方法 19.3.4 可变法(feasible method) 该法是中间距离法的推广,其类间距离的递推公式为

$$D_{kr}^2 = \frac{1-\beta}{2}[D_{kp}^2 + D_{kq}^2] + \beta D_{pq}^2, \quad \beta < 1.$$

可变法适用于$\{d_{ij}\}$为非度量尺度的情形.

方法 19.3.5 类平均法(group average method) 该法也是对最长和最短距离法的折衷,其类间距离定义为

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{i \in G_p, j \in G_q} d_{ij}^2,$$

其中 n_p 与 n_q 分别为 G_p 和 G_q 中点的个数. 设 $G_r = \{G_p, G_q\}$,G_k 为另一个类,则递推公式为

$$D_{kr}^2 = \frac{n_p}{n_r} D_{kp}^2 + \frac{n_q}{n_r} D_{kq}^2.$$

有人曾对 7 种层次聚类法做过试验,认为类平均法是效果最好的一种.

方法 19.3.6 可变类平均法(feasible group average method) 它是将类平均法的递推公式改为

$$D_{kr}^2 = \frac{n_p}{n_r}(1-\beta) D_{kp}^2 + \frac{n_q}{n_r}(1-\beta) D_{kq}^2 + \beta D_{pq}^2, \quad \beta < 1.$$

方法 19.3.7 重心法(gravity centre method) 以类重心间的距离为类间距离. 设 $\bar{x}_p$ 和 $\bar{x}_q$ 分别为类 G_p 和 G_q 的重心,则

$$D_{pq} = d_{\bar{x}_p \bar{x}_q}.$$

距离递推公式为

$$D_{kr}^2 = \frac{n_p}{n_r} D_{kp}^2 + \frac{n_q}{n_r} D_{kq}^2 - \frac{n_p n_q}{n_r^2} D_{pq}^2.$$

方法 19.3.8 沃德法(Ward's method) 此方法是沃德于 1963 年提出来的. 它借用了方差分析的思想,即要求同类样品之间的离差平方和应较小,而类间离差平方和应较大. 使用这种方法的前提是存在对 p 个变量的 n 个样品的观测数据(前几种方法不

受此限制，只需要有相似阵或相异阵即可). 设类 G_k 中有 n_k 个点，$x_{ik}, i=1,\cdots,n_k$，用 $\bar{x}_k$ 记 G_k 的重心(类样本均值). 则在 G_k 中的样品离差平方和为

$$S_k = \sum_{i=1}^{n_k}(x_{ik} - \bar{x}_k)'(x_{ik} - \bar{x}_k).$$

当将类 G_p 和 G_q 合为 G_r 时有 $S_r \geqslant S_p + S_q$，因此定义类间距离为

$$D_{pq}^2 = S_r - S_p - S_q.$$

因此将 D_{pq} 最小的两类合并相当于使类内离差平方和增加最小. 递推公式(按欧氏距离)为

$$D_{kr}^2 = \frac{n_k + n_p}{n_r + n_k}D_{kp}^2 + \frac{n_k + n_q}{n_r + n_k}D_{kq}^2 - \frac{n_k}{n_r + n_k}D_{pq}^2.$$

公式 19.3.9 统一的距离递推公式　以上几种层次聚类法的类间距离可用一个统一公式来表示：

$$D_{kr}^2 = \alpha_p D_{kp}^2 + \alpha_q D_{kq}^2 + \beta D_{pq}^2 + \gamma |D_{kp}^2 - D_{kq}^2|,$$

其中 $\alpha_p, \alpha_q, \beta, \gamma$ 为参数(见表 19.1)，取不同值时得到不同类间距离的递推公式.

表 19.1

方　法	α_p	α_q	β	γ
最短距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
最长距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
中间距离法	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
可变法	$\frac{1-\beta}{2}$	$\frac{1-\beta}{2}$	<1	0
类平均法	n_p/n_r	n_q/n_r	0	0
可变类平均法	$(1-\beta)n_p/n_r$	$(1-\beta)n_q/n_r$	<1	0
重心法	n_p/n_r	n_q/n_r	$-(n_p n_q)/n_r$	0
沃德法	$\frac{n_k+n_p}{n_k+n_r}$	$\frac{n_k+n_q}{n_k+n_r}$	$-\frac{n_k}{n_k+n_r}$	0

算法 19.3.10 算法概述 (1)规定客体之间的距离，计算客体距离阵，记作 $D_{(0)}$，这时每个客体算一类；(2)选 $D_{(0)}$的最小元素，设为 D_{pq}，将类 G_p 和 G_q 合并为一个新类，记为 $G_r=\{G_p,G_q\}$；(3)利用递推公式计算新类与其它类的距离 D_{rk}，将 $D_{(0)}$中的第 p，q 行和第 p,q 列划掉，代之以由 D_{rk}构成的新行新列，可得矩阵记作 $D_{(1)}$.

回到第(2)步，直至得到某个 $D_{(k)}$只有一行一列为止.

19.3.3 非层次聚类法

非层次聚类(nonhierarchical clustering)是一类聚类方法的统称，其共同特点是：先给定一个粗糙的初始分类，然后按照某种原则反复进行修改，直至分类较为合理为止. 非层次聚类法比之层次聚类法，其最主要的优点是占用计算机内存少，分类较准确.

在非层次聚类中，为了得到初始分类，通常要先选择一批**凝聚点**(condensation point)，让样品向凝聚点集中. 所谓凝聚点就是一批有代表性的点，是要形成类的中心. 选择凝聚点有以下的办法.

方法 19.3.11 均值法 将数据人为地分为 k 类，计算每一类的均值，将这些均值作为凝聚点.

方法 19.3.12 密度法 人为地定两个正数 $d_1<d_2$，以每个样品点为中心，以 d_1 为半径，到这个样品点的距离小于 d_1 的样品个数代表该中心样品点的密度. 首先选具有最大密度的样品点作为第一凝聚点，记为 z_1. 再选距第一凝聚点 z_1 远于 d_2 的具有最大密度的点作为第二凝聚点，记为 z_2. 再取距 z_1 和 z_2 均远于 d_2 且有最大密度的点作为第三凝聚点，记为 z_3. 如此反复，直到没有新的凝聚点产生为止.

在选定的凝聚点的基础上进行分类和修改的方法有很多种，下面介绍三种比较定型的方法：

方法 19.3.13 K-均值法(K-means method) (1) 人为地定

出分类数目 K，将所有样品任意地分成 K 类，计算各类的重心作为凝聚点.

(2) 将 n 个样品从头至尾输入一遍，每进入一个样品，将它归入距它最近的凝聚点所代表的类，重新计算类重心，以新的类重心作为凝聚点.

重复(2)，直至所得分类不再改变.

K-均值法计算简便，分类迅速，但由于人为地确定了类的个数 K，有时会影响分类效果.

方法 19.3.14 改进的 K-均值法 人为地定三个数 K,C,R.

(1) 取前 K 个样品作为凝聚点，计算两两之间的距离. 如果最小距离小于 C，则将相应的两凝聚点合并，用重心作为新的凝聚点，重复以上步骤，直至所有凝聚点之间的距离大于等于 C 为止.

(2) 将余下的 $n-K$ 个样品逐个进入，每进入一个样品就计算它与各凝聚点之间的距离. 如最小距离大于 R，则该样品作为新的凝聚点；如果最小距离小于等于 R，则将该样品归入最近凝聚点的那一类，重新计算该类重心，以此重心作为新的凝聚点. 重新验证凝聚点之间的距离，如有小于 C 的，则用(2)中的方法合并.

(3) 将 n 个样品从头至尾逐个输入，用第(2)步的方法归类，不同之处是：当某个样品从原来的类进入另一个类时，两类的重心都需要重新计算.

重复(3)，直至分类不发生变化为止.

这个方法中 K,C,R 的选择会影响分类的效果.

方法 19.3.15 等数据法(isodata method) 人为选择 7 个控制参数 NS，TN，TE，NT，TC，ND 及 IX.

(1) 输入一批凝聚点，或用某种方法产生一批凝聚点.

(2) 将样品归入最近凝聚点所代表的类，在 n 个样品没有全部归类之前凝聚点不变. 当 n 个样品全部归类后，重新计算每类重

心，作为新的凝聚点，重复这个步骤 NS 次或者至收敛为止.

（3）若某一类的样品数小于 TN，则该类样品不参加下面的计算.

（4）按如下的原则将类进行合并或分解：

1）如类的数目 $k \geqslant 2$ND，则要合并；

2）如类的数目 $k \leqslant \frac{1}{2}$ND，则要分解；

3）如类的数目 k 在上述界限之间，则当 k 是奇数时要分解，偶数时合并.

（5）计算各类重心，将它们作为凝聚点，回到第(2)步.

（6）重复(3)至(5)步 IX 次.

其中(4)步中所说的合并过程，重心需重新算过. 如最近的两类的距离小于 TC，则合并这两类，并重新计算这个新类与其它类之间的距离，这个过程最多进行NT 次.

在分解过程中，对某个类 G_i，如第 j 个变量的标准差 $\sigma_j^{(i)} >$ TEσ_j，其中 σ_j 为第 j 个变量的所有 n 个样品的标准差，则将 G_i 一分为二，以第 j 个变量在 G_i 中的均值来划线，均值以上的为一类，其余的为另一类. 计算这两个新类的重心，这两个重心之间的距离如大于 1.1TC，则这个分解是采纳的，否则不作分解.

19.4 多维标度法

多维标度法(multidimational scaling)，简称为 MDS，是一类统计分析方法的统称，产生于心理学度量分析，并在许多领域中得到了广泛的应用. MDS 的基本特点是由相似性或相异性数据出发，用低维空间中的**点结构**(configuration of points)来表示研究客体，从而揭示数据的潜在结构. 本节介绍几种得到广泛应用的 MDS 方法.

19.4.1 古典 MDS

古典 MDS(classical MDS)是最早出现的一种 MDS 方法,由塔格森(Torgerson)于 1950 年引进.

设 $\Delta=(\delta_{ij})$ 为一个 $n\times n$ 相异性矩阵,δ_{ij}表示客体 i 与客体 j 之间的相异性(19.3),古典 MDS 的目标是用某个 r 维欧氏空间中的 n 个点 $x_1,x_2,\cdots,x_n$ 来表示 n 个研究客体.记

$$d_{ij}=\|x_i-x_j\|=\left\{\sum_{k=1}^{r}(x_{ik}-x_{jk})^2\right\}^{\frac{1}{2}}, \qquad (19.23)$$

d_{ij}表示点 $x_i=(x_{i1},\cdots,x_{ir})$和 $x_j=(x_{j1},\cdots,x_{jr})$之间的欧氏距离.在古典 MDS 模型中将相异性数据近似看成是点之间的距离,即

$$\delta_{ij}\approx d_{ij}. \qquad (19.24)$$

方法 19.4.1 古典 MDS 的计算方法

(1) 令 $\boldsymbol{A}=(a_{ij})_{n\times n}$, $a_{ij}=\dfrac{-1}{2}\delta_{ij}^2$;

(2) 令 $\boldsymbol{B}=\boldsymbol{HAH}$,其中 $\boldsymbol{H}=\left(\boldsymbol{I}_n-\dfrac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n{}'\right)$;

$\boldsymbol{B}$ 称为**伪中心化内积阵**(pseudo centralized inner product matrix);

(3) 将 $\boldsymbol{B}$ 作特征分解,设 $\boldsymbol{B}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}}$,其中 $\boldsymbol{P}$ 为 $n\times n$ 正交阵,$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$,其中 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 为 B 的特征根.

(4) 设 $\lambda_r<0$,则在 r 维欧氏空间中,古典 MDS 没有解.否则($\lambda_r\geqslant0$),令 $\Lambda_r=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_r)$,$\boldsymbol{P}_r$ 为 $\boldsymbol{P}$ 的前 r 列构成的($n\times r$)矩阵,令

$$\underset{n\times r}{\boldsymbol{X}}=\boldsymbol{P}_r\boldsymbol{\Lambda}_r^{\frac{1}{2}}.$$

将 $\boldsymbol{X}$ 的每一行看成是 r 维欧氏空间中的点,$\boldsymbol{X}$ 即为古典 MDS 的解.

古典 MDS 依据的数学原理是:当 $\delta_{ij}=d_{ij}$时,在(2)中计算出

来的 **B** 矩阵一定是中心化内积矩阵，即 B 的元素 b_{ij} 满足：

$$b_{ij} = (x_i - \bar{x})(x_j - \bar{x})', \tag{19.25}$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 为结构的中心，而且此时定有

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r \geqslant 0, \quad \lambda_{r+1} = \cdots = \lambda_n = 0. \tag{19.26}$$

且

$$\lambda_j = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2, \quad j = 1, \cdots, r. \tag{19.27}$$

λ_j 为 n 个点在第 j 个坐标上的方差，λ_j 的大小可描述第 j 个坐标的重要性. 由 λ_j 的单调性，知第 1 坐标最重要，第 2 坐标次重要，…等等.

在(3)中，当 $\lambda_r<0$ 时，古典 MDS 在 r 维空间中没有解. 这时有两条路径可循. 第一条路径是事先不指定 r，当 **B** 分解以后，取 r 满足 $\lambda_r>0$，且

$$\sum_{j=1}^{r}\lambda_j \Big/ \sum_{j=1}^{n}|\lambda_j| \geqslant 0.85. \tag{19.28}$$

第二条路径是取常数 c 满足

$$c = -\sum_{i=r+1}^{n-1}\lambda_i/(n - r - 1), \tag{19.29}$$

c 称为**加性常数**(additive constant)，在(3)中令 $\Lambda_r=\mathrm{diag}(\lambda_1+c, \lambda_2+c, \cdots, \lambda_r+c)$，再进行计算即可，采用加性常数法可保证对 $r\leqslant n-1$ 总有解.

古典 MDS 的模型是近似将相异性看成点间距离，模型拟合的优度可用下述两个量来描述：

$$\begin{aligned} \alpha_1 &= \Big(\sum_{j=1}^{r}\lambda_j \Big/ \sum_{j=1}^{n}|\lambda_j|\Big) \times 100\%, \\ \alpha_2 &= \Big(\sum_{j=1}^{r}\lambda_j^2 \Big/ \sum_{j=1}^{n}\lambda_j^2\Big) \times 100\%, \end{aligned} \tag{19.30}$$

α_1 称为**百分变差**(percentage variation)；α_2 称为**平方百分变差**

(squared percentage variation).

当数据是相似性时,算法略有不同.记 $C=(c_{ij})$ 为相似阵.即 c_{ij} 表示客体 i 与 j 之间的相似性.在计算过程中,直接令伪中心化内积矩阵为

$$B = HCH \tag{19.31}$$

即可.以后的计算方法是相同的.

下面介绍一个古典 MDS 的例子.

例 19.4.2 数据是 10 个莫尔斯电码之间的相似性,列于下表.

表 19.2

电码	1	2	3	4	5	6	7	8	9	0
1	84	63	13	08	10	08	19	32	57	55
2	62	89	54	20	05	14	20	21	16	11
3	18	64	86	31	23	41	16	17	08	10
4	05	26	44	89	42	44	32	10	03	03
5	14	10	30	69	90	42	24	10	06	05
6	15	14	26	24	17	86	69	14	05	14
7	22	29	18	15	12	61	85	70	20	13
8	42	29	16	16	09	30	60	89	61	26
9	57	39	09	12	04	11	42	56	91	78
0	50	26	09	11	05	22	17	52	81	74

用古典 MDS 方法进行分析,得到 B 矩阵的特征根为

$$\lambda_1 = 187.4,\ \lambda_2 = 121.0,\ \lambda_3 = 95.4,$$
$$\lambda_4 = 55.4,\ \lambda_5 = 46.6,\ \lambda_6 = 31.5,\ \lambda_7 = 9.6,$$
$$\lambda_8 = 4.5,\ \lambda_9 = 0.0,\ \lambda_{10} = -4.1$$

假定要求一个二维解,则 B 的前两个特征向量(模分别为 λ_1,λ_2)为

(−4.2, −0.3, 3.7, 5.6, 5.4, 3.8, 0.9, −3.0, −6.2, −5.7)

(−3.2, −5.8, −4.3, −0.6, 0.0, 4.0, 5.5, 3.6, 0.6, 0.2)

用以上两个向量的对应坐标表示二维欧氏空间中的点可以得到下图 19.2.

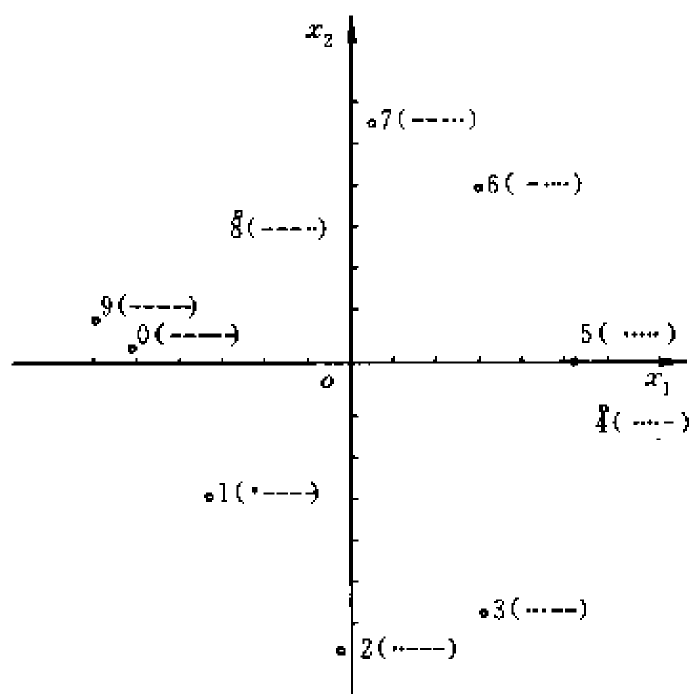


图　19.2

代表 1～0 的莫尔斯电码列于表 19.3.

表　19.3

数　字	电　码　点　　划
1	· — — — —
2	· · — — —
3	· · · — —
4	· · · · —
5	· · · · ·
6	— · · · ·
7	— — · · ·
8	— — — · ·
9	— — — — ·
0	— — — — —

根据上述的图和表，可以看出：第一个坐标度量"点"的个数("点"即莫尔斯电码中的·)，点多的码在轴的右端，点少的码在轴的左端. 第二个坐标度量电码的不均匀性. 以上两个因素构成区别莫尔斯电码的主要原因.

19.4.2 非度量 MDS

非度量 MDS 是谢帕德(Shepard)于 1962 年提出的一种 MDS 模型，并由克鲁斯卡尔(Kruskal)于 1964 年给出了一种有效的算法.

非度量 MDS 模型的基本特征是将相似性或相异性数据看成是点间距离的单调函数，以下统一用 δ_{ij}记客体 i 与客体 j 之间的相似性或相异性，d_{ij}为点 x_i 与点 x_j 之间的距离. 在非度量 MDS 模型中假定

$$\delta_{ij} \approx f(d_{ij}), \tag{19.32}$$

其中，当 δ_{ij}为相似性时 f 为单调降函数，当 δ_{ij}为相异性时 f 为单调增函数. 上述要求的实质是要求最后计算出来的点结构的点间距离与数据之间应有一种保序的关系. 即对相异性数据，应有

$$\delta_{ij} \leqslant \delta_{kl} \Rightarrow d_{ij} \leqslant d_{kl}. \tag{19.33}$$

对相似性数据应有

$$\delta_{ij} \leqslant \delta_{kl} \Rightarrow d_{ij} \geqslant d_{kl}. \tag{19.34}$$

在理想的情形下，(19.33)、(19.34)式可以满足，但在大多数实际情形下，上述单调关系只能在一定程度上得到满足，克鲁斯卡尔给出了上述单调性满足程度的度量，并通过最小化这种度量给出有效的算法.

定义 19.4.3 设 $\Delta=(\delta_{ij})_{n\times n}$为一个相异性矩阵，$\boldsymbol{X}_{n\times n}$为 r 维空间中 n 个点的坐标阵，d_{ij}为点 x_i 与 x_j 之间的距离，对 δ_{ij}的单调增函数 $\hat{d}_{ij}=f(\delta_{ij})$，定义

$$S_1(\boldsymbol{X},f)=\Big[\sum_{ij}(d_{ij}-\hat{d}_{ij})^2\Big/\sum_{ij}d_{ij}^2\Big]^{\frac{1}{2}},$$

$$S_2(\boldsymbol{X},f)=\Big[\sum_{ij}(d_{ij}-\hat{d}_{ij})^2\Big/\sum_{ij}(d_{ij}-\bar{d})^2\Big]^{\frac{1}{2}},$$

其中 $\bar{d}=\frac{1}{M}\sum_{ij}d_{ij}$，$M$ 为 δ_{ij}的有效个数，$S_1(\boldsymbol{X},f)$和 $S_2(\boldsymbol{X},f)$分别称为**克鲁斯卡尔应力** Ⅰ 和 Ⅱ.

在上述应力的定义中，当 δ_{ij} 为相似性时，$\hat{d}_{ij}$改为 δ_{ij}的单调降函数. 以下简记 $S_1(\boldsymbol{X},f)$和 $S_2(\boldsymbol{X},f)$为 S_1 和 S_2，克鲁斯卡尔建议用应力 S_1 或 S_2 来衡量拟合的优度，并给出通过最小化 S_1(或 S_2)来求解的算法. 这种算法的基本特点是一个二重迭代过程. 首先给定 $\boldsymbol{X}$ 的初始值，可考虑用古典解或随机初始值. 然后对给定的 $\boldsymbol{X}$，用单调回归方法得到一组$\{\hat{d}_{ij}\}$，然后固定这一组$\{\hat{d}_{ij}\}$再对 X 进行更新，更新的方法是一种经验的梯度法，在实际应用中非常有效. 上述过程反复进行，直到应力值降到某个指定的限以下，或者应力值不再发生显著变化为止.

19.4.3 个体差异标度法——INDSCAL

设有 m 个关于同一组客体的相异性(相似性)数据矩阵 Δ_1，$\Delta_2,\cdots,\Delta_m$，可能是来自不同的试验者或不同的试验场合(条件)，或不同的试验方法，以下统称这些不同的试验者(场合，方法)为主体 (subjects). 这样 m 个矩阵对应着 m 个主体. INDSCAL (individual difference scaling)是一种 MDS 模型，其目的不仅是要分析客体的结构，而且进一步要分析主体之间的差异，称为**个体差异标度法**.

在 INDSCAL 模型中，仍然假定 n 个客体可用 r 维空间中的 n 个点 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 来表示. 但是对不同的主体，点间距离的定义是不同的，设 $d_{ij,k}$ 为第 k 个主体关于客体 i 与客体 j 之间的距离，其表达式为

$$d_{ij},k = \left\{ \sum_{t=1}^{r} w_{kt}(x_{it} - x_{jt})^2 \right\}^{\frac{1}{2}}. \qquad (19.35)$$

d_{ij},k 相当于对第 t 个坐标加权 $\sqrt{w_{kt}}$ 后计算的欧氏距离，权 $\sqrt{w_{kt}}$ 代表第 k 个主体对第 t 个坐标的重视程度

根据式 19.35 知，对第 k 个主体来说，点 i 与点 j 的内积是 $\sum_{t=1}^{r} w_{kt} x_{it} x_{jt}$. 对每个数据阵 $\boldsymbol{\Delta}_k$，利用节 19.4.1 中计算伪中心化内积阵的方法计算 $\boldsymbol{B}_k = (b_{ij},k)$. INDSCAL 模型假定

$$b_{ij},k \approx \sum_{t=1}^{r} w_{kt} x_{it} x_{jt}. \qquad (19.36)$$

为求解权和点结构，采用最小化离差平方和方法，即令

$$R = \sum_{i,j,k} \left(b_{ij},k - w_{kt} \sum_{t=1}^{r} x_{it} x_{jt} \right)^2 \qquad (19.37)$$

达到最小，为最小化 R，通常采用的算法是**交替最小二乘法**(alternative least squares). 即将 R 改写为

$$\tilde{R} = \sum_{i,j,k} \left(b_{ij},k - w_{kt} \sum_{t=1}^{r} x_{it} y_{jt} \right)^2. \qquad (19.38)$$

在 $\tilde{R}$ 中包含了三组变量：w，x 和 y. 当固定其中两组时，对第三组变量来说，最小化 $\tilde{R}$ 是一个普通的最小二乘法问题. 所谓交替最小二乘法，就是每次固定两组变量，对第三组变量用最小二乘法最小化 $\tilde{R}$，整个过程可用交替最小二乘法示意图表示(图 19.3).

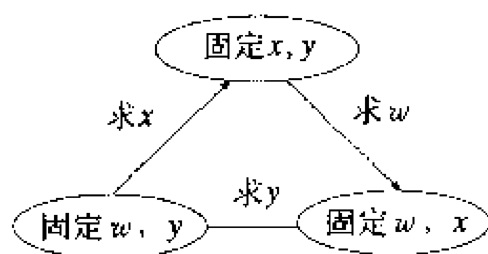


图 19.3

上述示意图表示交替最小二乘法的一个循环过程. 上述循环

过程一直进行到 $\tilde{R}$ 没有显著改变为止.

在 $\tilde{R}$ 中每当 x 和 y 改变一个尺度时,即令 $x^*=cx, y^*=cy$ 时,令 $w^*=c^{-2}w$,则 $\tilde{R}$ 的值不变.因此为保证解的唯一性,每当求出一个 x 时,要将其标准化,即令

$$\sum_{i=1}^{n} x_{it} = 0,\ \sum_{i=1}^{n} x_{it}^2 = 1,\quad t = 1,\cdots,r. \tag{19.39}$$

对 y 也有上述要求,当循环结束时要令 $x=y$,且令 $w=|w|$(因为 w 中可能有负值,而负权在 INDSCAL 模型中是没有意义的).

循环开始的初始值可取 x 为随机结构或者将 $\boldsymbol{B}_t$ 作加权平均后求古典 MDS 解,然后令 $y=x$.

MDS 方法的种类很多,例如还可以对非对称数据建立非对称模型,或处理长方形相似阵的模型等.

20 抽样调查方法

20.1 概念与符号

20.1.1 概述

抽样调查方法常用于人口调查、能源调查、社会经济调查、森林林木估积、草原及农田估产、野生生物数量估计等方面.抽样调查在调查研究中有极大的实用价值.合理地抽样既可获得可靠的高精度的信息,又可大大地节约调查的人力、物力、财力、时间.特别当信息有很强的时间性时,旷日持久的全面调查将只能获得陈旧的信息,而失去调查的价值.

方法 20.1.1 符号

(1) Π_N——N 个有明确编号的个体 $o_1, o_2, o_3, \cdots, o_N$ 所组成的**总体**.

(2) N——被调查全体对象的**总数**,是已知的正整数.

(3) Y_i——总体 Π_N 中第 i 个个体 O_i 所对应的某一**数量指标**,$i=1,2,\cdots N$.

(4) n——抽样的个数,它是从有 N 个个体的总体中用各种抽样办法抽取的 n 个个体.

(5) y_i——抽取的第 i 个个体所对应的数量指标,$i=1,2,\cdots,n$.($y_1, y_2, \cdots, y_n$ 是 $Y_1, Y_2, \cdots, Y_N$ 的一个部分).

待估计的量(总体数字特征)

(6) $\overline{Y} = \dfrac{1}{N}\sum\limits_{i=1}^{N} Y_i$ ——**总体平均值**.

(7) $\widetilde{Y} = \sum_{i=1}^{N} Y_i$ ——**总值**，即所调查的全体对象的数量指标之和.

(8) $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$ ——**总体方差**，

或
$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2.$$

(9) $C^2 = S^2/\overline{Y}^2$ ——**变异系数**.

抽样调查内容：

(1) 制定调查方案：包括抽样实施的方法，以及根据事先给出的误差标准，决定抽样的最小个数 n.

(2) 由 $y_1, y_2, \cdots, y_n$ 估计总体的数字特征，并研究估计的误差.

例 20.1.2 一块 100km^2 的地方，是血吸虫的中间宿主钉螺繁殖之地，今要调查钉螺总量，以便开展防治血吸虫病的工作. 为此将这 100km^2 的地域分成 1m^2 一块共 10^8 个小块，调查其中的 1000 个小块，计算出每个小块中挖出的钉螺数. 得 $y_1, y_2, \cdots, y_{1000}$. 此例中 $N = 10^8$, $n = 1000$, 1000km^2 的钉螺总数 $\widetilde{Y} = Y_1 + Y_2 + \cdots + Y_{10^8}$ 是待估计的量，要根据 $y_1, y_2, \cdots, y_{1000}$ 来估计 $\widetilde{Y}$.

20.1.2 抽样方法

方法 20.1.3 随机抽样法(random sampling) 在 Π_N 的 N 个个体中机会均等地抽取第一个样，然后在剩下的 $N-1$ 个个体中机会均等地抽取第二个样，……最后在所剩的 $N-(n-1)$ 个个体中机会均等地抽取第 n 个样.

方法 20.1.4 分层抽样法(hierarchical sampling) 将 Π_N 先分成 k 组：$\Pi_{N_1}, \Pi_{N_2}, \cdots, \Pi_{N_k}$，称 Π_{N_i} 为第 i 层. 把每一层 Π_{N_i} 看成一

个小总体，在 Π_{N_i} 中随机取一组大小为 n 的样本，$i=1,2,\cdots,k$. 这 k 组样本合成 Π_N 的分层样本. 当 Π_N 中的个体有明显差异时，则应将相近的个体归为一组. 如此将 Π_N 分成 k 组，采用分层抽样法取样.

方法 20.1.5 二阶抽样法(two-step sampling) 先将 Π_N 分成 k 组：$\Pi_{N_1},\Pi_{N_2},\cdots,\Pi_{N_k}$，这些组称为 Π_N 的第一性抽样单位，将它们看成 k 个个体进行抽样，例如抽到 $\Pi_{N_1},\Pi_{N_4},\Pi_{N_7}$ 组，然后再把 $\Pi_{N_1},\Pi_{N_4},\Pi_{N_7}$ 作为第二性抽样单位，在 $\Pi_{N_1},\Pi_{N_4},\Pi_{N_7}$ 中分别抽样.

方法 20.1.6 多阶抽样法(multiple-step sampling) 设想 20.1.5 中第二性抽样单位仍不是 Π_N 的个体，而是一小组，对它再做第三性抽样，余类推.

方法 20.1.7 集团抽样法(group sampling) 将 Π_N 分成 k 组：$\Pi_{N_1},\Pi_{N_2},\cdots,\Pi_{N_k}$，对这 k 个组进行抽样，对入样的 Π_{N_i} 不再个体进行抽样，而是将其中的个体逐个观察，第 i 组称第 i 个集团.

方法 20.1.8 系统抽样法(system-sampling) 又称机械抽样法. 选一正整数 k，将 Π_N 中的个体逐个排列如下：

$$
\begin{aligned}
&1,2,3,\cdots,k,\\
&k+1,k+2,k+3,\cdots,2k,\\
&2k+1,2k+2,2k+3,\cdots,3k,\\
&\cdots\cdots,\\
&\text{直至 } N \text{ 为止.}
\end{aligned}
$$

对号码 $1,2,\cdots,k$，做随机抽样(常常只抽一个)，若 i 入样，则上述 i 号码的那一列号码 $i,k+i,2k+i,\cdots$ 的个体全入样.

20.2 随机抽样法

20.2.1 随机抽样法

随机抽样法是等概地从总体 Π_N 中无放回地抽取 n 个样，具体地实现，可将 Π_N 中的 N 个个体标以号码 1，2，…，N. 然后利用随机数表抽出 n 个不同的数码，或用计算机产生伪随机数来抽取 n 个不同的数码. 用随机数表抽取的方法如下：设总体 Π_N 的个体数 $N=345$. 要抽取 $n=15$ 个样. 由于 $N=345$ 是三位数，所以从随机数表中任取三列所构成的三位数码中，依次取出不同的三位数码，当数码在 001～345 之间时，则该号码入样，当数码在 401～745 之间时，则将该数减去 400 后的号码入样，其余的 000，346～

表 20.1 随机数表

65	54	73	88	44	76	68	47	93	11	14	95	72	94	14
95	84	67	58	37	62	18	03	23	61	60	88	44	62	99
05	63	05	42	44	63	44	78	98	09	25	58	00	57	12
36	05	60	21	12	26	61	99	62	44	83	09	78	74	84
53	45	44	36	44	78	74	09	25	58	00	57	12	38	46
05	60	21	13	26	61	99	62	44	76	73	55	63	86	00
58	22	57	86	27	63	43	45	60	74	16	72	73	79	50
79	59	69	50	72	47	26	95	60	88	44	83	09	79	49
69	39	80	45	69	92	93	38	12	57	00	58	16	23	54
56	32	36	22	58	69	50	72	47	26	61	99	87	90	83
77	93	81	16	61	05	97	78	84	43	59	76	16	76	91
29	63	99	16	65	60	82	94	19	25	98	28	24	57	05
61	10	39	10	58	48	81	77	60	31	45	17	09	62	16
34	31	36	56	98	31	26	20	03	27	63	09	56	36	16
35	16	11	17	56	31	58	25	87	90	05	29	87	77	34

400,746～999 不要. 当某号码已入样,而再次碰到该号码时,则只算一次. 例如,用表 20.1 所示随机数表取样. 则可利用 8,9,10 列构成三位数. 第一个数为 844,舍去. 第二个数为 837,也舍去. 第三个数为 244,则 244 号入样. 第四个数为 112,112 号也入样. 第 5 个数为 644,减去 400 为 244,由于 244 号已入样,故该数也舍去. 如此继续可得 326,227,72,169,258,261,265,58,298 号入样. 8,9,10 列随机数已取完,但入样号码仍不足 15 个,此时可转到另三列继续取,如利用 1,2,3 列继续得 255,56,360,134 入样. 取满 15 个即停止.

20.2.2 简单估值法

设 $y_1,y_2,\cdots,y_n$ 是来自总体 Π_N 的随机样本. $\bar{Y},\hat{Y},\sigma^2,S^2,C^2$(见 20.1.1)为待估的量. 记

$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ ——样本均值,

$\hat{y}=N\bar{y}$,

$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$ ——样本方差,

$\sigma_{\bar{y}}^2=E(\bar{y}-\bar{Y})^2$——用 $\bar{y}$ 估计 $\bar{Y}$ 时的均方误差,

$C_{\bar{y}}^2=\frac{E(\bar{y}-\bar{Y})^2}{\bar{Y}^2}=\frac{\sigma_{\bar{y}}^2}{\bar{Y}^2}$,

其中 $\sigma_{\bar{y}}^2,C_{\bar{y}}^2$ 也是待估的量.

定理 20.2.1 给出了 $\bar{Y},S^2,\sigma_{\bar{y}}^2,C_Y,C$ 的无偏估计.

定理 20.2.1 设 $y_1,y_2,\cdots,y_n$ 是来自总体 $Y_1,Y_2,\cdots,Y_N$ 的随机样本,则

(1) $E(\bar{y})=\bar{Y}\quad(E(\hat{y})=E(\hat{Y}))$.

(2) $\sigma_{\bar{y}}^2=\frac{1-f}{n}S^2\quad(C_{\bar{y}}^2=\frac{1-f}{n}\frac{S^2}{\bar{Y}}=\frac{1-f}{n}C^2)$,

其中,$f=\frac{n}{N}$.

(3) $E(S^2)=S^2 \quad (E(s_y^2)=\sigma_y^2)$，其中 $S_y^2=\frac{1-f}{n}s^2$.

此定理告诉我们，用 $\bar{y}$ 估 $\bar{Y}$（$\hat{y}$ 估 $\hat{Y}$）；用 s^2 估 S^2；用 s_y^2估 σ_y^2，这些估计是无偏的.

方法 20.2.2　$\bar{Y}$ 及 $\hat{Y}$ 的区间估计　当 n 相当大时，$\bar{y}$ 接近 $N(\bar{Y},\sigma_y^2)$分布. 因而对于给定的置信度 $1-\alpha$ $(0\leqslant\alpha\leqslant 1)$，查 $N(0,1)$ 表，可得 $z_{\frac{\alpha}{2}}$.

$$
\begin{aligned}
1-\alpha &= P\left\{\left|\frac{\bar{y}-\bar{Y}}{\sigma_y}\right|<z_{\frac{\alpha}{2}}\right\} \\
&= P\left\{\bar{y}-z_{\frac{\alpha}{2}}\sigma_y^2<\bar{Y}<\bar{y}+z_{\frac{\alpha}{2}}\sigma_y^2\right\} \\
&= P\left\{\bar{y}-z_{\frac{\alpha}{2}}\sigma_y^2 N<\hat{Y}<\bar{y}+z_{\frac{\alpha}{2}}\sigma_y^2 N\right\}
\end{aligned}
$$

在实际工作中以 σ_y^2的无偏估计 $s_y=\sqrt{s_y^2}$ 代替 σ_y，即可得到 $\bar{Y}$ 与 $\hat{Y}$ 的区间估计. 其中 $S_y^2=\frac{1-f}{n}S^2$.

方法 20.2.3　部分估计(partial estimate)　设总体 Π_N 中有一"子体"$\Pi_{N'}$，它的大小 N'为未知，但 Π_N 中任一个体是否属于 $\Pi_{N'}$ 有明确的判别法. 现在要调查这"子体"$\Pi_{N'}$ 中的个体的数量指标之和及 $\Pi_{N'}$ 中个体的数目 N'. 设"子体"(用其数量指标表示)为 $\Pi_{N'}=\{Z_1,Z_2,\cdots,Z_{N'}\}$，今令 $\Pi_N-\Pi_{N'}$ 中的个体(即 Π_N 中不属于 $\Pi_{N'}$ 的个体)数量指标取作 0，则 $\Pi_N=\{Y_1,Y_2,\cdots,Y_N\}=\{Z_1,Z_2,\cdots,Z_{N'},\underbrace{0,0,\cdots,0}_{N-N'\text{个}}\}$.

调查对象为 $\hat{Y}=\sum_{i=1}^{N}Y_i=\sum_{i=1}^{N'}Z_i=\hat{Z}$，及 N'. 调查所取样本为 $(y_1,y_2,\cdots,y_n)=(z_1,z_2,\cdots,z_{n'},\underbrace{0,0,\cdots,0}_{n-n'\text{个}})$.

记 $p=\frac{n'}{n}$，则用 Np 估计 N'，用 $\frac{N}{n}\sum_{i=1}^{n'}z_i$ 估计 $\hat{Z}$.

方法 20.2.4 决定样本容量的方法

(1) 若要使 $\sigma_y=\sqrt{\sigma_y^2}\leqslant\delta$($\delta$ 事先给定),则要求

$$n\geqslant\frac{NS^2}{N\delta^2+S^2}.$$

(2) 若要使 $P\{|\bar{y}-\bar{Y}|\leqslant d\}=1-\alpha$ (α,d 事先给定),则要求

$$n\geqslant\frac{NS^2}{N\left(\frac{d}{z_{\frac{\alpha}{2}}}\right)^2+S^2},$$

其中 $z_{\frac{\alpha}{2}}$ 为标准正态分布的上$\frac{\alpha}{2}$分位数.

(3) 若要使 $C_{\bar{y}}<\nu$ $\left(C_{\bar{y}}^2=\frac{E(\bar{y}-\bar{Y})^2}{\bar{Y}^2}\right.$,即要求相对精度高于指定的指标 $\left.\nu\right)$,则要求

$$n\geqslant\frac{NC^2}{N\nu^2+C^2}.$$

(4) 若要使 $P\left\{\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right|<h\right\}=1-\alpha$. ($\alpha,h$ 事先指定),则要求

$$n\geqslant\frac{NC^2}{N\left(\frac{h}{z_{\frac{\alpha}{2}}}\right)^2+C^2}.$$

注:(1)~(4)公式中的 S^2,C^2 是未知量,在实际工作中,可根据过去的资料,或先进行少量抽样,预估出 S^2 或 C^2,从而粗略地确定样本额 n.

(5) 考虑调查费用决定样本量之方法举例.

例 20.2.5 设调查费用为 C_0+C_1n,C_0 为基本调查费用,C_1 为每调查一个样所需费用.又设以 $\bar{y}$ 估 $\bar{Y}$,由于误差 $|\bar{y}-\bar{Y}|$ 所造成的损失为 $AE(\bar{y}-\bar{Y})^2=A\left(\frac{1}{n}-\frac{1}{N}\right)S^2$,则要使总耗费最小,即 $C_0+C_1n+A\left(\frac{1}{n}-\frac{1}{N}\right)S^2$ 最小.应取

$$n=\sqrt{\frac{\Lambda S^2}{C_1}}.$$

其中 Λ 是将 $E(\bar{y}-\bar{Y})^2$ 折合成钱数的因子.

例 20.2.6 设调查费用为 $C_0+C_1 n$. 误差 $|\bar{y}-\bar{Y}|$ 所造成的损失为 $\Lambda E|\bar{y}-\bar{Y}|$. 为使总耗费 $C_0+C_1 n+\Lambda E|\bar{y}-\bar{Y}|$ 最小,此时可认为 $\bar{y}-\bar{Y}$ 近似服从 $N\left(0,\frac{\sigma^2}{n}\right)$ 分布. $E|\bar{y}-\bar{Y}|=\frac{\sqrt{2}}{\sqrt{\pi}}\frac{\sigma}{\sqrt{n}}\int_0^{\infty}x\exp\left[-\frac{x^2}{2\sigma^2/n}\right]\mathrm{d}x=\sqrt{\frac{2}{\pi}}\frac{\sigma}{\sqrt{n}}$. 要使 $C_0+C_1 n+\Lambda\sqrt{\frac{2}{\pi}}\frac{\sigma}{\sqrt{n}}$ 最小,应取

$$n=\left(\frac{\Lambda^2\sigma^2}{2\pi C_1^2}\right)^{1/3}.$$

实际工作中,可由历史资料或少量抽样预估出 S^2 或 σ^2.

20.2.3 比估值法

设总体中每个个体有两个指标 X,Y. 于是有

$$\begin{pmatrix}X_1 & X_2 & \cdots & X_N\\ Y_1 & Y_2 & \cdots & Y_N\end{pmatrix}.$$

通过计算有

$$\begin{matrix}\bar{X} & \hat{X} & \sigma_X^2 & S_X^2 & C_X^2\\ \bar{Y} & \hat{Y} & \sigma_Y^2 & S_Y^2 & C_Y^2\end{matrix}$$

相关系数

$$\rho=\frac{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{S_X^2 S_Y^2}}.$$

我们调查的目标是,通过样本 $\begin{pmatrix}x_1 & x_2 & \cdots & x_n\\ y_1 & y_2 & \cdots & y_n\end{pmatrix}$ 来调查 $R=\frac{\bar{Y}}{\bar{X}}=$

$\dfrac{\widetilde{Y}}{\widetilde{X}}$. 当 $\overline{X}$ 是已知时，这时估计 R 与估计 $\overline{Y}$ 是同一问题；有时 $\overline{X}$ 是未知的. 例如，调查某城市 1987 年的人口（Y）相对于 1955 年人口（X）的比率，1955 年的人口 X 是已知的.

又例如调查一片庄稼病株所占的比例，个体（人工单位面积）的两个指标是

$$X —— 株数, \qquad Y —— 病株数,$$

调查目标为 $R=\dfrac{\widetilde{Y}}{\widetilde{X}}=\left(=\dfrac{\overline{Y}}{\overline{X}}\right)$. 这种情况 X 则是未知的.

定理 20.2.7 设总体为 $\begin{bmatrix} X_1 & X_2 & \cdots & X_N \\ Y_1 & Y_2 & \cdots & Y_N \end{bmatrix}$，$X_i>0$，$\overline{Y}>0$. 令 $R=\dfrac{\overline{Y}}{\overline{X}}$，样本为 $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}$. 记 $r=\dfrac{\overline{y}}{\overline{x}}$，则有

(1) $E(r)=R+O(\dfrac{1}{n})$.

(2) $$E(r-R)^2=\frac{1-f}{n}\frac{1}{\overline{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n^{3/2}}\right)=O\left(\frac{1}{n}\right).$$

(3) $$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n}\right).$$

(4) $$E\left[\frac{1}{\overline{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]=\frac{1}{\overline{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n^{1/2}}\right).$$

由定理 20.2.7 可知以下方法.

方法 20.2.8 (1) 用 r 估 R，其估计为有偏估计，其偏 $E(r)-R=O\left(\dfrac{1}{n}\right)$. 均方误差为 $E(r-R)^2=O\left(\dfrac{1}{n}\right)$. 由定理 20.2.7 之

(2),(4)知, $E(r-R)^2 \doteq \frac{1-f}{n}\frac{1}{\overline{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$,近似地可用$\frac{1-f}{n}\frac{1}{\overline{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$估计$E(r-R)^2$.

(2) 如果$\overline{X}$已知,可用$r\overline{X}$估$\overline{Y}$,其估计为有偏估计.均方偏差为$E(r\overline{X}-Y)^2 \doteq \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$.近似地可用$\frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$估计$E(r\overline{X}-\overline{Y})^2$.

定义 20.2.9 称相对均方偏差

$$e=\frac{\sigma_{SE}^2-\sigma_{RE}^2}{\sigma_{RE}^2}=\frac{C_{SE}^2-C_{RE}^2}{C_{RE}^2}=\frac{2\rho C_X C_Y-C_X^2}{C_Y^2-2\rho C_X C_Y+C_X^2}$$

为比估值法对简单估值法的效率增量.

其中$\sigma_{RE}^2=E(r\overline{X}-\overline{Y})^2$表示用$r\overline{X}$估计$\overline{Y}$时所产生的均方偏差.

$\sigma_{SE}^2=\sigma_{\overline{y}}^2=E(\overline{y}-\overline{Y})^2$表示在简单估值法中用$\overline{y}$估$\overline{Y}$时所产生的平均平方偏差.

定理 20.2.10 比估值法优于简单估值法,即效率增量$e>0$的充要条件是$\rho>\frac{C_X}{2C_y}$.

由定理可知,当使用辅助指标X时,X的变异系数C_X^2愈小愈好,而X与调查对象Y的相关系数ρ则愈大愈好.

20.3 分层抽样法

20.3.1 分层简单估值法

设总体Π_N分为k层

$$\Pi_N=\Pi_{N_1}\cup\Pi_{N_2}\cup\cdots\cup\Pi_{N_k},$$

其中$\Pi_{N_\nu}=\{Y_{\nu_1},Y_{\nu_2},\cdots,Y_{\nu_{N_\nu}}\}$, $\nu=1,2,\cdots,k$; $N=N_1+N_2+\cdots$

$+N_k$.

对每层定义参数：

$$\overline{Y}_\nu = \frac{1}{N_\nu}\sum_{i=1}^{N_\nu} Y_{\nu_i}, \quad \nu = 1,2,\cdots,k;$$

$$\widetilde{Y}_\nu = \sum_{i=1}^{N_\nu} Y_{\nu_i}, \quad \nu = 1,2,\cdots,k;$$

$$\sigma_\nu^2 = \frac{1}{N_\nu}\sum_{i=1}^{N_\nu}(Y_{\nu_i} - \overline{Y}_\nu)^2, \quad \nu = 1,2,\cdots,k;$$

$$S_\nu^2 = \frac{1}{N_\nu - 1}\sum_{i=1}^{N_\nu}(Y_{\nu_i} - \overline{Y}_\nu)^2, \quad \nu = 1,2,\cdots,k;$$

$$C_\nu^2 = S_\nu^2/\overline{Y}_\nu^2, \quad \nu = 1,2,\cdots,k.$$

各符号意义与节 20.1.1 完全相同，只是多了标志层号的下标 ν.
记

$$W_\nu = \frac{N_\nu}{N}, \quad \nu = 1,\cdots,k;$$

$$\overline{Y} = \sum_{\nu=1}^{k} W_\nu \overline{Y}_\nu = \frac{1}{N}\sum_{\nu=1}^{k}\widetilde{Y}_\nu \text{——总体 } \Pi_N \text{ 的指标均值;}$$

$$\widetilde{Y} = N\overline{Y} = \sum_{\nu=1}^{k}\widetilde{Y}_\nu \text{——总体 } \Pi_N \text{ 的指标和.}$$

调查对象是 $\overline{Y}$ 或 $\widetilde{Y}$.

样本符号：

n_ν——在第 ν 层的抽样数，$\nu=1,2,\cdots,k$；

$y_{\nu_1}, y_{\nu_2}, \cdots, y_{\nu_{n_\nu}}$——是在第 ν 层中抽取的样本指标，$\nu=1,2,\cdots,k$.

统计量：

$$\overline{y}_\nu = \frac{1}{n_\nu}\sum_{i=1}^{n_\nu} y_{\nu_i} \text{——第 } \nu \text{ 层抽样指标的均值;}$$

$$\widetilde{y}_\nu = N_\nu \overline{y}_\nu;$$

$$S_{y_\nu}^2 = \frac{1}{n_\nu - 1}\sum_{i=1}^{n_\nu}(y_{\nu_i} - \bar{y}_\nu)^2 \text{——第 } \nu \text{ 层样本方差.}$$

并记

$$\bar{y}_{SSE} = \sum_{\nu=1}^{k} W_\nu \bar{y}_\nu = \frac{1}{N}\sum_{\nu=1}^{k} \tilde{y}_{\nu i}.$$

$$S_{SSE}^2 = \sum_{\nu=1}^{k} W_\nu S_{y_\nu}^2;$$

$$\sigma_{SSE}^2 = E(\bar{y}_{SSE} - \bar{Y})^2.$$

定理 20.3.1

(1) $E(\bar{y}_{SSE}) = \bar{Y}$,

(2) $\sigma_{SSE}^2 = \sum_{\nu=1}^{k} W_\nu^2 \frac{1-f_\nu}{n_\nu} S_\nu^2$,其中 $f_\nu = n_\nu / N_\nu$.

(3) $E(S_{SSE}^2) = \sigma_{SSE}^2$.

方法 20.3.2 估计方法 用 $\bar{y}_{SSE}$估 $\bar{Y}$ 是无偏估计.

用 $N\bar{y}_{SSE}$估 $\tilde{Y}$ 是无偏估计.

用 S_{SSE}^2估 σ_{SSE}^2是无偏估计.

20.3.2 按比例分配样本额的分层抽样法

按比例分配样本额是指取

$$n_\nu = \frac{N_\nu}{N} n = W_\nu n, \quad n = n_1 + n_2 + \cdots + n_k.$$

其估计方法与 20.3.1 估计方法相同,只是在相应的公式中 n_ν 的地方用 $n_\nu = W_\nu n$ 代之.

按比例分配样本额分层抽样方法与简单不分层随机抽样法之比较即 SSE(按比例)与 SE 法之比较. SSE 法对 SE 法的效率增量为

$$e = \frac{\sigma_{SE}^2 - \sigma_{SSE}^2}{\sigma_{SSE}^2}$$

$$=\frac{\frac{N}{N-1}\sum_{\nu=1}^{k}W_\nu(\overline{Y}_\nu-\overline{Y})^2-\sum_{\nu=1}^{k}\left(W_\nu-\frac{N_\nu-1}{N-1}\right)S_\nu^2}{\sum_{\nu=1}^{k}W_\nu S_\nu^2}$$

$$\doteq\frac{\sum_{\nu=1}^{k}W_\nu(\overline{Y}_\nu-\overline{Y})^2}{\sum_{\nu=1}^{k}W_\nu S_\nu}.$$

由此可见，层平均的差别愈大，即层间差异愈大，而各层内方差 S_ν^2 愈小，则效率增量愈大. 所以总体 Π_N 中的个体间有明显差异时，则应将性质相近的划归为同一层，将 Π_N 分成若干层，采用分层抽样法.

分层抽样法(按比例分配样本额)决定最小样本额.

(1) 要使 $\sigma_{\mathrm{SSE}}^2=E(\overline{y}_{\mathrm{SSE}}-\overline{Y})^2\leqslant\delta$，$\delta$ 为事先指定的正数，要求

$$n\geqslant\frac{1}{\frac{\delta}{\sum_{\nu=1}^{k}W_\nu S_\nu^2}+\frac{1}{N}}.$$

(2) 要使 $C_{\mathrm{SSE}}^2\leqslant r$，$r$ 为事先指定的正数 $C_{\mathrm{SSE}}^2=\frac{\sigma_{\mathrm{SSE}}^2}{\overline{Y}^2}$，要求

$$n\geqslant\frac{1}{\frac{r\overline{Y}^2}{\sum_{\nu=1}^{k}W_\nu S_\nu^2}+\frac{1}{N}}.$$

(1)，(2)公式中有未知参数 $\overline{Y}$ 及 S_ν^2，这些量可用其它办法先做出预估.

20.3.3 样本额的理想分配

按比例分配样本额考虑了各层中个体数的不同这一差异性，而未考虑各层方差之不同，一般说来方差小的层样本可少抽一些.

因此有一个更合理的分配样本额的问题.同时考虑到各层中个体数目的不同又考虑到各层中方差不同这两个因素来分配样本额称之为理想样本分配额,它由下面的定理来决定.

定理 20.3.3 理想样本分配额为

$$n_\nu=\frac{W_\nu S_\nu}{\frac{1}{n}\sum_{\nu=1}^{k}W_\nu S_\nu}=n\frac{W_\nu S_\nu}{\sum_{\nu=1}^{k}W_\nu S_\nu},\quad \nu=1,2,\cdots,k.$$

因此要想算出理想的样本分配额 n_ν,必须知道各层的方差 S_ν. 而 S_ν 是未知的.

方法 20.3.4 S_ν 的预估法

(1) 预查法(pretest method) 先在每一层中少量取一序样,各层的取样量分别为 $m_1,m_2,\cdots,m_k$. 记这些序样为 $(z_{\nu_1},z_{\nu_2},\cdots,z_{\nu_{m_\nu}})$, $\nu=1,2,\cdots,k$. 以 $\frac{1}{m_{\nu-1}}\sum_{i=1}^{m_\nu}(z_{\nu_i}-\bar{z}_\nu)^2$ 近似代替 S_ν^2. 再求得理想样本额 n_ν.

(2) 察往法 若总体过去被调查过,则现在按过去的 S_ν 估计分配样本额.

(3) 辅助指标法(supplementary variabl method) 总体 Π_N 中的个体,除要调查的指标 Y 外,往往另有一个"规模指标",这种"规模指标"是度量个体的规模大小的. 例如调查全国人口出生率,调查个体为县,则各县的现有人口就是一个"规模指标". 假定分层按"规模指标"的大小. 如调查人口出生率时,将全国各县按拥有人口的多少分层. 此时可假定各层 S_ν 与 $\overline{X}_\nu$ 成正比. 从而可得理想分配样本额.

20.3.4 分层比估值法

设总体中的每个个体有两个指标 $\begin{pmatrix}X_i\\Y_i\end{pmatrix}$,将总体分成 k 层. $\Pi_{N_1},\Pi_{N_2},\cdots,\Pi_{N_k}$.

设总体的第 ν 层是

$$\begin{bmatrix} X_{\nu_1} \\ Y_{\nu_1} \end{bmatrix}, \begin{bmatrix} X_{\nu_2} \\ Y_{\nu_2} \end{bmatrix}, \cdots, \begin{bmatrix} X_{\nu_{N_\nu}} \\ Y_{\nu_{N_\nu}} \end{bmatrix}, \quad \nu = 1,2,\cdots,k.$$

按照 20.2.3 符号的记法，各层相应的参数为

$$\begin{array}{lllll} \overline{X}_\nu & \widetilde{X}_\nu & \sigma^2_{X_\nu} & S^2_{X_\nu} & C^2_{X_\nu} \\ \overline{Y}_\nu & \widetilde{Y}_\nu & \sigma^2_{Y_\nu} & S^2_{Y_\nu} & C^2_{Y_\nu} \qquad \nu = 1,2,\cdots,k. \\ & R_\nu & & \rho_\nu & \end{array}$$

仅增加了标志层的下标.

总体的指标为 $X,\widetilde{X},\overline{Y},\widetilde{Y}$.

调查对象是 $\overline{Y}$ 或 $\widetilde{Y}$.

$Y=\sum_{\nu=1}^{k} W_\nu \overline{Y}_\nu$，其中 $W_\nu=\dfrac{N_\nu}{N}$.

设第 ν 层的样本是

$$\begin{bmatrix} X_{\nu_1} \\ Y_{\nu_1} \end{bmatrix}, \begin{bmatrix} X_{\nu_2} \\ Y_{\nu_2} \end{bmatrix}, \cdots, \begin{bmatrix} X_{\nu_{n_\nu}} \\ Y_{\nu_{n_\nu}} \end{bmatrix}, \quad \nu = 1,2,\cdots,k.$$

相应的统计量为 $\overline{x}_\nu,\overline{y}_\nu,r_\nu=\dfrac{\overline{y}_\nu}{\overline{x}_\nu}$.

方法 20.3.5 估计方法 假定辅助指标 $\overline{X}_\nu(\nu=1,2,\cdots,k)$已知，则在每一层中用比估值法.

用 $r_\nu\overline{X}_\nu$ 估 $\overline{Y}_\nu,\nu=1,2,\cdots,k$. 用 $\overline{y}_{SRE}$估 $\overline{Y}$，用 S^2_{SRE}估 σ^2_{SRE}，

其中 $\overline{y}_{SRE} = \sum_{\nu=1}^{k} W_\nu r_\nu \overline{X}_\nu$，

$$S^2_{SRE} = \sum W^2_\nu \frac{1-f_\nu}{n_\nu}\frac{1}{n_\nu-1}\Big(\sum_{i=1}^{n_\nu} y^2_{\nu_i} - 2r_\nu \sum_{i=1}^{n_\nu} x_{\nu_i}y_{\nu_i} + r^2_\nu \sum_{i=1}^{n_\nu} x^2_{\nu_i}\Big).$$

估计的性质是：

用 $r_\nu\overline{X}_\nu$ 估 $\overline{Y}_\nu$，其均方偏差为

$$E(r_\nu\overline{X}_\nu - \overline{Y}_\nu)^2 \doteq \frac{1-f_\nu}{n_\nu}[S^2_{Y_\nu} - 2R_\nu\rho_\nu S_{Y_\nu}S_{X_\nu} + R^2_\nu S^2_{X_\nu}];$$

用 $\bar{y}_{SRE}$ 估 $\bar{Y}$，其均方偏差为

$$\sigma_{SRE}^2 = E(\bar{y}_{SRE} - \bar{Y})^2 \doteq \sum_{\nu=1}^{k} W_\nu^2 \frac{1-f_\nu}{n_\nu}[S_{Y_\nu}^2 - 2R_\nu \rho_\nu S_{Y_\nu} S_{X_\nu} + R_\nu^2 S_{X_\nu}^2];$$

用 S_{SRE}^2 估 σ_{SRE}^2，为近似无偏的，$E(S_{RES}^2) \doteq \sigma_{SRE}^2$；用 $\bar{y}_{SRE}$ 估计 $\bar{Y}$，估计是近似无偏的，$E(\bar{y}_{SRE}) \doteq \bar{Y}$.

通常情况下，使用与 Y 关系密切的辅助指标 X，采用分层比估值法比用简单分层估值法能得到更好的估计.

20.4 二阶抽样法

凡作调查，必有一批完全明确的最小单位，它们组成总体. 前几章所讨论的抽样法皆认为最小单位可直接编号，再进行随机抽样. 但是实际工作中往往遇到最小单位不易直接编号的情形. 例如要调查全市各小学中学生的学习情况、健康情况等各项指标. 由于小学生数量极大，他们的情况由他们所在的小学掌握. 因此可以不把全市小学生统一编号，而把全市小学编号，作为第一性抽样单位，在所抽到的学校中再将学生编号，作为第二性抽样单位. 即采用二阶抽样法. 用二阶抽样法，由于被抽个体只集中于第一阶抽样的若干小学中，因而具有节力，节时等优点.

20.4.1 二阶抽样问题的一般提法

设总体 Π_N 分为 k 组：

$\Pi_N = \{\Pi_{N_1}, \Pi_{N_2}, \Pi_{N_3}, \cdots, \Pi_{N_K}\}$ 对每个组 $\Pi_{N_i}(i=1,2,\cdots,k)$ 都有一个调查指标 G_i（G_i 可以是 $\bar{Y}_i$ 或 $\tilde{Y}_i, \cdots$），组内调查的目的是估计 G_i，而总的调查目的是要估计

$$G = G_1 + G_2 + \cdots + G_k.$$

因为可能抽到任一组，从而可能在任一组内进行抽样调查，所以事先应对每一个组拟定一个抽样计划，并选定对组指标的估计方法

(例如在节 20.2,20.3 中所介绍的某种估计方法).

将上述的有关量,给以适当的符号,见表 20.2.

假定所选之 g_i 是 G_i 的无偏估计量,$S^2_{g_i}$是 $\sigma^2_{g_i}$的无偏估计量.

表 20.2

组	Π_N			
	Π_{N_1}	Π_{N_2}	…	Π_{N_k}
组　指　标	G_1	G_2	…	G_k
拟定的组内抽样法	某法	某法	…	某法
拟定的估计量	g_1	g_2	…	g_k
估计量的均方偏差	$\sigma^2_{g_1}$	$\sigma^2_{g_2}$	…	$\sigma^2_{g_k}$
均方偏差的估计量	$S^2_{g_1}$	$S^2_{g_2}$	…	$S^2_{g_k}$

总体指标 G 的估计方法不仅依赖于组内的抽样法(即第二阶抽样)而且依赖于如何抽组(即第一阶抽样).

方法 20.4.1 第一阶抽样法(抽组法)

(1) 随机抽组法:等概无放回地抽组 K 次.

(2) 有返回$(p_1,p_2,\cdots,p_K)$抽组法:有返回地抽组 K 次,每次抽到 Π_{N_i}组的概率为 p_i.

20.4.2 二阶抽样之估值法

抽组法是随机抽组法时之估值法.

定理 20.4.2 设用随机抽组法在 K 组中抽到 k 个组为 $\Pi_{N_{\theta_1}}$,$\Pi_{N_{\theta_2}}$,$\cdots$,$\Pi_{N_{\theta_k}}$,则有

(1) $g=\dfrac{K}{k}\sum\limits_{i=1}^{k}g_{\theta_i}$ 是 G 的无偏估计量,即有 $E(g)=G$.

(2) g 的均方偏差为 $\sigma^2_g=E(g-G)^2=\dfrac{K^2}{k}V_1+\dfrac{K^2}{k}(1-$

$f)V_2$，其中 $V_1 = \frac{1}{K}\sum_{i=1}^{K}\sigma_{g_i}^2, f = \frac{k}{K}, V_2 = \frac{1}{K-1}\sum_{i=1}^{K}\left(G_i - \frac{G}{K}\right)^2$.

(3) $\hat{V}_1 = \frac{1}{k}\sum_{i=1}^{k}S_{g_{\theta_2}}^2$ 是 V_1 的无偏估计量.

(4) $\hat{V}_2 = \frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i} - \frac{g}{K}\right)^2 - \frac{1}{k}\sum_{i=1}^{k}S_{g_{\theta_i}}^2$ 是 V_2 的无偏估计量.

(5) 若 $\sigma_{g_{\theta_i}}^2$ 不存在无偏估计量 $S_{g_{\theta_i}}^2$，则以 $\hat{\sigma}_g^2 = \frac{K}{k}(1-f)\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i} - \frac{g}{K}\right)^2$ 来估计 σ_g^2，这个估计是有偏的，但其偏与 σ_g^2 相比小于 k/K.

抽组法是有返回($p_1, p_2, \cdots, p_k$)时之估计法.

定理 20.4.3 设用有返回($p_1, p_2, \cdots, p_k$)抽组法，得到 k 个组为

$$\Pi_{N_{\theta_1}}, \Pi_{N_{\theta_2}}, \cdots, \Pi_{N_{\theta_k}},$$

则有

(1) $g' = \frac{1}{k}\sum_{i=1}^{k}\frac{g_{\theta_i}}{p_{\theta_i}} = \frac{1}{k}\sum_{i=1}^{k}z_i$ 是 G 的无偏估计量，其中

$$z_i = g_{\theta_i}/p_{\theta_i}, \quad i = 1, 2, \cdots, k.$$

(2) g' 的均方偏差 $\sigma_{g'}^2 = E(g' - G)^2$.

$$\sigma_g^2 = \frac{1}{k}\sum_{i=1}^{k}\frac{\sigma_{g_i}^2}{p_i} + \frac{1}{k}\sum_{i=1}^{k}p_i\left(\frac{G_i}{p_i} - G\right)^2.$$

(3) $\hat{\sigma}_{g_i}^2 = \frac{1}{k(k-1)}\sum_{i=1}^{k}(z_i - \bar{z})^2$ 是 $\sigma_{g'}^2$ 的无偏估计.

其中 $\bar{z} = \frac{1}{k}\sum_{i=1}^{k}z_i$.

上述定理 20.4.2 及 20.4.3 之估值法，不论组内的抽样方案如何都适用. 如果组内抽样是采用随机抽样法，组指标 $G_i = \bar{Y}_i$. 采

用二阶抽样法的估值法在下面介绍.

20.4.3 组内为随机抽样时的二阶抽样估值法

组指标 $$G_i = \widetilde{Y}_i = \sum_{i=1}^{N_i} Y_{ij}.$$

总的调查对象 $$G = \widetilde{Y} = \widetilde{Y}_1 + \widetilde{Y}_2 + \cdots + \widetilde{Y}_k.$$

在组 Π_{N_i} 内采用随机抽样法(参看方法 20.1.3).

(1) G_i 的无偏估计量为

$$g_i = \widetilde{y}_i = \frac{N_i}{n_i}\sum_{i=1}^{n_i} y_i.$$

(2) g_i 的均方偏差 $\sigma_{g_i}^2 = E(g_i - G_i)^2$,

$$\sigma_{g_i}^2 = N_i^2 \frac{1 - f_i}{n_i} S_i^2,$$

其中 $f_i = \dfrac{n_i}{N_i}$,

$$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(Y_{ij} - \overline{Y}_i)^2,$$

$$\overline{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}.$$

(3) $\sigma_{g_i}^2$ 的无偏估计量为

$$S_{g_i}^2 = N_i^2 \frac{1 - f_i}{n_i} s_i^2,$$

其中 $s_i^2 = \dfrac{1}{n_i - 1}\displaystyle\sum_{i=1}^{n_i}(y_{ij} - \overline{y}_i)^2$, $\overline{y}_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{n_i} y_{ij}$.

20.4.4 随机抽组法

抽组法是随机抽组法(简记为 TSE),而且组内抽样是随机抽样法时,总指标 $\overline{Y}$,$\widetilde{Y}$ 的估计方法为

(1) 用 $\tilde{y}_{TSE}=\frac{K}{k}\sum_{i=1}^{k}\tilde{y}_{\theta_i}$ 估计 $\tilde{Y}=\tilde{Y}_1+\tilde{Y}_2+\cdots+\tilde{Y}_k$，此估计是无偏的，即 $E(\tilde{y}_{TSE})=\tilde{Y}$.

(2) $\sigma^2_{\tilde{y}_{TSE}}=E(\tilde{y}_{TSE}-\tilde{Y})^2$ 为

$$\sigma^2_{\tilde{y}_{TSE}}=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2.$$

其中 $f=\frac{k}{K}$，

$$V_1=\frac{1}{K}\sum_{i=1}^{K}\sigma^2_{g_i}=\frac{1}{K}\sum_{i=1}^{K}\frac{N_i^2}{n_i}(1-f_i)S_i^2,$$

$$V_2=\frac{1}{K-1}\sum_{i=1}^{K}\left(G_i-\frac{G}{K}\right)^2=\frac{1}{K-1}\sum_{i=1}^{K}\left(\tilde{Y}_i-\frac{\tilde{Y}}{K}\right)^2,$$

$\hat{V}_1=\frac{1}{k}\sum_{i=1}^{k}S^2_{g_{\theta_i}}=\frac{1}{k}\sum_{i=1}^{k}\frac{N^2_{\theta_i}}{n_{\theta_i}}\left(1-f_{\theta_i}\right)S^2_{\theta_i}$ 是 V_1 的无偏估计量.

$$\hat{V}_2=\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2-\frac{1}{k}\sum_{i=1}^{k}S^2_{g_{\theta_i}}$$

$$=\frac{1}{k-1}\sum_{i=1}^{k}\left(\tilde{y}_{\theta_i}-\frac{1}{k}\sum_{i=1}^{k}\tilde{y}_{\theta_i}\right)^2-\frac{1}{k}\sum_{i=1}^{k}\frac{N^2_{\theta_i}}{n_{\theta_i}}\left(1-f_{\theta_i}\right)S^2_{\theta_i}$$

是 V_2 的 无偏估计量.

20.4.5 返回抽组法

抽组法是有返回($p_1,p_2,\cdots,p_k$)法(简记 TPE)，且组内抽样是随机抽样法时，总指标 $\overline{Y}$,$\tilde{Y}$ 的估计为

(1) 用 $\tilde{y}_{TPE}=\frac{1}{k}(z_1+z_2+\cdots+z_k)$

$$=\frac{1}{k}\left[\frac{\tilde{y}_{\theta_1}}{p_{\theta_1}}+\frac{\tilde{y}_{\theta_2}}{p_{\theta_2}}+\cdots+\frac{\tilde{y}_{\theta_k}}{p_{\theta_k}}\right].$$

来估计 $\tilde{Y}$，其估计为无偏的，即

$$E(\tilde{y}_{TPE})=\tilde{Y}$$

(2) 用 $\tilde{y}_{TPE}$估计 $\tilde{Y}$ 的均方偏差为

$$\sigma^2_{\tilde{y}_{\mathrm{TPE}}} = E(\tilde{y}_{\mathrm{TPE}} - \tilde{Y})^2$$
$$= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{p_i}\frac{N_i^2(1-f_i)}{n_i}S_i^2 + \frac{1}{k}\sum_{i=1}^{k}p_i\left(\frac{\tilde{Y}_i}{p_i} - \tilde{Y}\right)^2.$$

$$\sigma^2_{\tilde{y}_{\mathrm{TPE}}} = \frac{1}{k(k-1)}\sum_{i=1}^{k}(z_i - \bar{z})^2.$$

其中 $z_i = \dfrac{\tilde{y}_{\theta_i}}{p_{\theta_i}}$, $\bar{z} = \dfrac{1}{k}\sum_{i=1}^{k} z_i$. 且 $E(\hat{\sigma}^2_{\tilde{y}_{\mathrm{TPE}}}) = \sigma^2_{\tilde{y}_{\mathrm{TPE}}}$.

当 $N_1 = N_2 = \cdots = N_k \equiv N_0, n_1 = n_2 = \cdots = n_k = n_0$ 时，$\tilde{Y}$ 及 $\bar{Y}$ 的估计可用下面方法得到：

（3）当 $N_1 = N_2 = \cdots \equiv N_0, n_1 = n_2 = \cdots = n_0$ 时，用 $\tilde{y}_{\mathrm{TPE}} = \dfrac{z_1 + z_2 + \cdots + z_k}{k}$ 估计 $\tilde{Y}$，其中 $z_i = \dfrac{N_0\bar{y}_{\theta_i}}{p_{\theta_i}}$. 此估计是无偏的. 其均方偏差为 $\sigma^2_{\tilde{y}_{\mathrm{TPE}}} = \dfrac{N_0^2(1-f_0)}{kn_0}\sum_{i=1}^{k}\dfrac{S_i^2}{p_i} + \dfrac{N_0^2}{k}\sum_{i=1}^{k}p_i\left(\dfrac{\bar{Y}_i}{p_i} - K\bar{Y}\right)^2$,

$$\hat{\sigma}^2_{\tilde{y}_{\mathrm{TPE}}} = \frac{1}{k(k-1)}\sum_{i=1}^{k}(z_i - \bar{z})^2, \text{其中 } z_i = \frac{N_0\tilde{y}_{\theta_i}}{p_{\theta_i}}.$$

（4）当 $N_1 = N_2 = \cdots = N_K \equiv N_0, n_1 = n_2 = \cdots = n_k = n_0$ 时，用 $\bar{y}_{\mathrm{TPE}} = \dfrac{1}{k}\left(\dfrac{\bar{y}_{\theta_1}}{Kp_{\theta_1}} + \dfrac{\bar{y}_{\theta_2}}{Kp_{\theta_2}} + \cdots + \dfrac{\bar{y}_{\theta_k}}{Kp_{\theta_k}}\right)$ 估计 $\bar{Y}$，此估计是无偏的. 其均方偏差为

$$\sigma^2_{\bar{y}_{\mathrm{TPE}}} = \frac{(1-f_0)}{kn_0K^2}\sum_{i=1}^{K}\frac{S_i^2}{p_i} + \frac{1}{k}\sum_{i=1}^{K}p_i\left(\frac{\bar{Y}_i}{Kp_i} - \bar{Y}\right)^2,$$

$$\hat{\sigma}^2_{\bar{y}_{\mathrm{TPE}}} = \frac{1}{k(k-1)}\sum_{i=1}^{k}\left(\frac{\bar{y}_{\theta_i}}{Kp_{\theta_i}} - \frac{1}{kK}\sum_{i=1}^{k}\frac{\bar{y}_{\theta_i}}{p_{\theta_i}}\right)^2.$$

21 随机过程的一般理论

21.1 引言与基本概念

21.1.1 引言与例

事物的变化过程可分为两大类. 在每个固定时刻 t,变化的结果,一类是确定的,这结果可用 t 的某个确定性函数描述,例如 $\sin t$;另一类是随机的,即以某种可能性出现多个(有限或无限多)结果之一,这结果可用与 t 有关的某个随机变量描述,例如泊松型变量 $X_t \sim P(\lambda t)$ 为参数 $\lambda, t>0$. 前一类变化过程称为**确定性过程**(deterministic process),而后一类即**随机过程**(stochastic process). 显然,前者可视为后者的特例. 随着研究的深入,许多原先用确定性方法处理的,都逐渐为随机方法所代替,即深化为随机过程的研究.

随机过程理论产生于20世纪初期,首先由于物理学、生物学、通讯与控制、以及管理科学等方面的需要而逐步发展起来的. 现在,随机过程的理论与方法几乎深入到每个自然与社会科学的分支.

例 21.1.1 直线上的随机游动(random walk) 设一粒子从零点出发,每次随机地向左或向右移动一个长度单位,概率分别为 $q\,(>0)$, $p\,(=1-q>0)$. 以 X_n 记该粒子 n 次移动后的位置,则 $\{X_n, n=0,1,2,\cdots\}$ 是一个随机过程.

例 21.1.2 以 X_t 记至 t 时刻止某电话交换台接通的电话次数,则 $\{X_t, t\geqslant 0\}$ 是一随机过程,若以 ζ_t 记至 t 时刻止,向此交换台的电话呼叫次数(要求接通电话),则 $\{\zeta_t, t\geqslant 0\}$ 也是随机过程.

例 21.1.3 以 X_n 表示未来第 n 年某地地下水位的高度(或年平均气温等),则 $\{X_n,n=1,2,\cdots\}$ 是随机过程.

例 21.1.4 布朗运动(Brownian motion) 英国植物学家 R. Brown 注意到悬浮在液面的微小粒子,不间断地进行着"杂乱无章"的运动. 粒子的这种运动,是由于大量相互独立的分子随机碰撞形成的. 如用 $\zeta_t\triangleq(X_t,Y_t)$ 表该粒子在时刻 t 的位置,则 $\{\zeta_t,t\geqslant 0\}$ 是一随机过程,通常它是二维布朗运动.

21.1.2 随机方法与确定性方法

为研究事物变化过程或系统性,建立数学模型时,有两种方法:随机性方法与确定性方法. 这两种方法中所涉及的数学方程也许形式类似,但它们要回答的问题的本质及所得结果的解决,却有重大差异.

例 21.1.5 随机性方法与确定性方法比较 设某细菌群体的个数在 $(t,t+\Delta t]$ 时段内只能增加,增加的数量与 t 时细菌数成正比. 开始时 $(t=0)$ 细菌数为 $x_0>0$.

(1) 确定性方法 设时刻 t 的细菌数为 $x(t)$,依设 $\Delta x(t)=\lambda x(t)\Delta t,\lambda>0$,令 $\Delta t\to 0$ 得 $\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=\lambda x(t)$. 代入初值解得 $x(t)=x_0\mathrm{e}^{\lambda t}$,这里 $x(t)$ 是实值连续函数.

(2) 随机性方法 设 t 时细菌数为一随机变量 X_t. $(t,t+\Delta t]$ 内增加的细菌数与 Δt 有关而与 t 无关,则 $X_t=x$ 条件下 $X_{t+\Delta t}$ 变为 $x+1$ 个的概率为

$$P(X_{t+\Delta t}=x+1\mid X_t=x)=\lambda x\Delta t+o(\Delta t).$$

又设在 $(t,t+\Delta t]$ 内增加不少于两个细菌的概率为 $o(\Delta t)$,则

$$P(X_{t+\Delta t}=x)=(1-\lambda x\Delta t)P(X_t=x)$$
$$+\lambda(x-1)\Delta tP(X_t=x-1)+o(\Delta t).$$

令 $\Delta t\to 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}P(X_t = x) = -\lambda x P(X_t = x) + \lambda(x-1)P(X_t = x-1)$$

初始条件 $P(X_0=x)=1$，当 $x=x_0$；　$x=0$，其它 x.

解得 $P(X_t=x)=C_{x-1}^{x-x_0}\mathrm{e}^{-\lambda x_0 t}(1-\mathrm{e}^{-\lambda t})^{x-x_0}, x \geqslant x_0$. 与确定性方法不同，在固定时刻 t，这里不是给定一个确定的细菌数而是对每数对 (t,x)，$x \geqslant x_0$ 给定一个概率 $P(X_t=x)\in[0,1]$. 确定性模型可认为 $P(X_t=x_0\mathrm{e}^{\lambda t})=1$. 又注意

$$EX_t \triangleq \sum_{x=0}^{\infty} xP(X_t = x) = x_0\mathrm{e}^{\lambda t}.$$

正是确定性问题的解（当然并非所有系统均如此），可见随机方法的解还考虑了群体增长的起伏.

21.1.3　随机过程的基本概念

1. 随机过程定义

定义 21.1.6　设给定一参数集 T 和可测空间 $(E,\mathscr{B})$，若对每个 $t\in T$，有一个定义在同一概率空间 $(\Omega,\mathscr{F},P)$ 上的 $\mathscr{B}$ 可测函数 $X_t(\omega)$，$\omega\in\Omega$ 与之对应，则称可测函数族 $X_T \triangleq \{X_t(\omega), t\in T\}$ 为定义在 $(\Omega,\mathscr{F},P)$ 上取值于 $(E,\mathscr{B})$ 中的**随机过程**，称 $(E,\mathscr{B})$ 为 X 的**状态空间**(state space)或**相空间**(phase space)，E 中元素称为**状态**(state). 当 $(E,\mathscr{B})$ 已明确指明时，简称 X_T 为随机过程或过程，而 E 为状态空间. 又称 T 为**参数空间**(parameter space).

定义 21.1.7　当 E 为实数(复数)集，$\mathscr{B}$ 为 E 上博雷尔可测集全体时的随机过程 X_T 称为**实(复)值随机过程**. $(E,\mathscr{B})=\langle R_n,\mathscr{B}_n\rangle$ 时称 X_T 为 **n 维实值随机过程**. 当 T 中元素 t 是向量时，称 X_T 为**随机场**(stochastic field)，或多参数过程.

由定义 21.1.7 可知，实(复)值随机过程 $X_T(\omega)$，对每一固定 $t\in T$，$X_t(\omega)$ 是 $(\Omega,\mathscr{F})$ 上实(复)值随机变量；对每一固定 $\omega\in\Omega$，$X_t(\omega)$ 是 T 上实(复)值函数. 这时也称它是过程 X_T 的一个**样本函**

数(sample function)．或**轨道**(trajectory)或**路径**(path)或**现实**(realization)．

注 21.1.8 约定除另有声明外，本卷以下 T 恒指参数集，随机过程恒为实值的．也常记 X_T 为 $X_t, t\in T$，$T=[0,\infty)$ 时也写作 $X_t, t\geqslant 0$．$X_t(\omega)$ 视为 $T\times\Omega$ 上的二元函数时常记为 $X(t,\omega)$，在不致混淆时常略去 ω，写为 $X_t, X(t)$ 等，也常以其它大写字母 Y, Z 等及希腊字母 ζ, η 等记之．

2．独立性与有限维分布函数族

定义 21.1.9 如对任意的 n 及不同的 $t_i\in T, i=1,2,\cdots,n$．n 个随机变量 $X_{t_i}, i=1,2,\cdots,n$ 为相互独立的(当 T 有限时，n 不超过 T 的元素数)．称过程 X_T 为**独立的**(independent，如对任意的 m 及不同的 $\lambda_j\in\Lambda, j=1,2,\cdots,m$ 有 $P(X_{\lambda_j}(\omega)\in B_{\lambda_j}, j=1,2,\cdots,m)=\prod_{j=1}^{m}P(X_t\lambda_j(\omega)\in B_{\lambda_j})$ 称**过程族** $X_{T_\lambda}, \lambda\in\Lambda(\subset R_1)$ 是**相互独立的**．这里 $B_{\lambda_j}\in\mathscr{B}^{T_{\lambda_j}}$，后者是乘积空间 $E^{T_{\lambda_j}}$ 中的乘积 σ 代数．

定义 21.1.10 设参数集 T 已知，对给定的任意 n 个 $t_i\in T$，$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 为 n 元分布函数，则称

$$F\triangleq\{F_{t_1,\cdots,t_n}(x_1,\cdots,x_n): t_i\in T,$$
$$x_i\in R_1, i=1,2,\cdots,n, n>0\}$$

为**有限维分布函数族**(family of finite dimensional distribution functions)，如 $F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 是 $X_{t_1},\cdots,X_{t_n}$ 的联合分布，则称 F 为(实值)过程 X_T 的有限维分布函数族．

定理 21.1.11 科尔莫戈罗夫存在定理(Kolmogorov existence theorem) 设 F 为有限维分布函数族，如对 $1,\cdots,n$ 的任一排列 $i_1,\cdots,i_n$，有 $F_{t_1\cdots t_n}(x_1\cdots x_n)=F_{t_{i_1}\cdots t_{i_n}}(x_{i_1},\cdots,x_{i_n})$；又如 $m<n$，则 $F_{t_1\cdots t_m}(x_1,\cdots,x_m)=\lim\limits_{x_j\to\infty, j=m+1,\cdots,n}F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$，则存在概率空间$(\Omega,\mathscr{F},P)$及其上的随机过程 X_T 使 X_T 的有穷维分

布族 F_X 与 F 重合.

定理中 F 满足的条件称为**相容性条件**(consistency condition).

3. 均值函数与协方差函数

随机过程 X_T 由它的有限维分布函数族 F_X 决定.给定 F_X 就能对 X_T 作全面的研究.但在许多实际问题中,要知道 F_X 是很困难的,常常用统计的方法,求出 X_T 的某些特征,从而对 X_T 的某些特性做讨论,这些特性往往正是我们要研究的实际问题的焦点,又易于运算.有时,这些特性能决定全部过程,像概率论中知道数学期望和方差就能完全决定一个正态分布的随机变量的概率分布.因此,对它们的研究很重要.

这些特征,最有用的是均值函数与协方差函数.

定义 21.1.12 设 X_T 为复值过程,且下面涉及的数学期望存在.称 $m(t)\triangleq EX_t, t\in T$ 为 X_T 的**均值函数**(mean function),而称

$$\begin{aligned}\Gamma(s,t)&\triangleq E(X_S-m(s))\overline{(X_t-m(t))}\\&=EX_S\overline{X}_t-m(s)\overline{m(t)},\quad s,t\in T\end{aligned}$$

为 X_T 的**协方差函数**(covariance function).称 $V(t)\triangleq\Gamma(t,t)=E|X_t-m(t)|^2, t\in T$ 为 X_T 的**方差函数**(variance function). $\sqrt{V(t)}$ 为**均方根差**(mean square root variance)函数或**标准差**(standard deviation)函数.也记 $V(t)$ 为 $\mathrm{var}(t)$ 或 $D(t)$.称 $R(s,t)\triangleq E\ X_S\overline{X}_t, s,t\in T$ 为 X_T 的(自)**相关函数**(auto-correlation function).

当 X_T 为实值时,$\Gamma(s,t)=E(X_S-m(s))(X_t-m(t))$,而当 X_T 均值函数恒为零时,$\Gamma(s,t)=R(s,t)$.

对均值函数与相关函数的研究是相关理论(correlation theory)的主要内容.

定义 21.1.13 设 X_T 和 Y_T 是同一概率空间上的两个复值过

程，且 $E|X_t|^2<\infty$，$E|Y_t|^2<\infty$，$t\in T$. 它们的均值函数分别为 $m_X(t)$，$m_Y(t)$，$t\in T$. 则称 $\Gamma_{XY}(s,t)\triangleq E(X_s-m_X(s)\overline{(Y_t-m_Y(t))})$，$s,t\in T$，为 X_T 与 Y_T 的**互协方差函数**(cross covariance function)，而 $R_{XY}(s,t)\triangleq EX_s\overline{Y_t}$，$s,t\in T$ 为它们的**互相关函数**(cross correlation function). 如 $\Gamma_{XY}(s,t)\equiv 0$，$\forall\ s,t\in T$，则称 X_T 与 Y_T 不相关.

易见 $\Gamma_{XY}(s,t)=R_{XY}(s,t)-m_X(s)\overline{m_Y(t)}$ 故 X_T 与 Y_T 不相关等价于 $R_{XY}(s,t)=m_X(s)\overline{m_Y(t)}$，$\forall\ s,t\in T$. 由定理 4.4.8 及例 4.4 知，两过程独立一定不相关，反之不一定成立.

定理 21.1.14 设 X_T 为多值过程，$E|X_t|^2<\infty$，$t\in T$，则 $\Gamma(s,t)$ 有如下性质

(1) 复对称性 $\Gamma(s,t)=\overline{\Gamma(t,s)}$.

(2) 非负定性 对每一 n 及任意的 $t_i\in T$ 与复数 λ_i，$i=1,2,\cdots,n$ 有

$$\sum_{i,j=1}^{n}\Gamma(t_i,t_j)\lambda_i\overline{\lambda_j}\geqslant 0.$$

21.2 随机过程的分类

21.2.1 按参数集与状态空间分类

按随机过程 $X_T=\{X_t,t\in T\}$ 的参数集 $T(\subset R_1)$ 与状态空间 E 是否可列，可分为 4 类：

(1) T,E 都可列.

(2) T 不可列，E 可列.

(3) T 可列，E 不可列.

(4) T,E 都不可列.

当 T 可列时，也称 X_T 为**离散参数随机过程**(stochastic process with discrete parameter)，或**随机序列**(random sequence)

或**时间序列**(time series). 当 T 连续(即 $T=R_1$ 或为 R_1 中一有限和无限区间)时,称 X_T 为**连续参数**(continuous parameter)**过程**. E 可列(有限)时,也称 X_T 为**可列(有限)过程**(denumerable (finite) process),或**离散状态过程**(discrete state process). E 连续的 X_T 为**连续状态过程**(continuous state process),例 21.1.1 给出的过程是离散参数离散状态过程属于上述分类(1),例 21.1.2~21.1.4 分别是(2)~(4)中的过程.

21.2.2 按过程的性质特点分类

按过程的性质特点,即 X_t 间的概率关系,也可对过程进行分类. 按此分类研究得较多的过程类型,有二阶矩过程,马尔可夫过程、独立增量过程及鞅. 当参数 t 是矢量时,X_T 是随机场. 这些类型是有交叉的,不像 21.2.1 节的分类互不相容.

本段恒设 $T=[0,\infty)$.

1. 二阶矩过程

定义 21.2.1 如过程 $X_T=\{X_t,t\geqslant 0\}$,对每 $t\geqslant 0$,X_t 的方差都存在,则称它是**二阶矩过程**(second order process)或**有限方差过程**(finite variance process).

由定理 4.3.2 及柯西-施瓦茨(Cauchy-Schwartz)不等式,一个二阶矩过程的均值函数与协方差函数都存在. 研究二阶矩过程特性,主要是研究协方差函数特性,并以此为基础,研究过程的解析性质(连续性,均方可微与可积等). 协方差函数的傅里叶变换称为谱密度函数,因此二阶矩的宽平稳过程特性又可用谱分析方法进行讨论.

二阶矩过程的详细讨论见第 22 章,其重要子类有正态过程,(宽)平稳过程(第 23 章)与正交增量过程.

定义 21.2.2 谐振过程 设 $X_t=\zeta\cos\theta t+\eta\sin\theta t$,$0\leqslant t\leqslant 1$(即 $T=[0,1]$),其中,随机变量 ζ 与 η 独立同分布 $N(0,\sigma^2)$,θ 为常

数. 称 X_T 为谐振过程，易知它是二阶矩过程. 据多元正态分布性质：$(X_{t_1},X_{t_2},\cdots,X_{t_n})$ 是 n 维正态的充要条件是对任意不全为零的实数 $a_1,a_2,\cdots,a_n$ 和 $\sum_1^n a_iX_{t_i}=\left(\sum_1^n a_i\cos\theta t_i\right)\zeta+\left(\sum_1^n a_i\sin\theta t_i\right)\eta$ 是一元正态的，由于 ζ,η 独立且正态，后一条件满足，从而 X_T 为正态过程(22.3.1). 易求得

$$m(t)=0,\quad \Gamma(s,t)=\sigma^2\cos\theta(s-t).$$

于是 X_T 也是零均值(宽)平稳过程.

2. 马尔可夫过程

粗略地说，马尔可夫过程是具有无后效性的随机过程. 所谓无后效性，是指已知过程现在的条件下，它将来的统计规律与过去无关. 一般工程系统中的随机过程都是马尔可夫过程.

定义 21.2.3 设 X_T 是以可测空间 $(E,\mathscr{B})$ 为状态空间，称 X_T 是马尔可夫过程. 如 $\forall A\in\mathscr{B}$ 以及 $t_1<t_2<\cdots<t_m<t_n\leqslant t$，有如下性质(称马尔可夫性).

$$P(X_t\in A|X_{t_1},X_{t_2},\cdots,X_{t_n})=P(X_t\in A|X_{t_n}),\text{a.s}$$

其中 $P(\cdot|X_{t_1},\cdots,X_{t_j})$ 指条件概率变量，它是 $\sigma(X_{t_1},\cdots,X_{t_j})$ 可测的. 特别，E 可列时，称 X_T 为**离散参数马尔可夫过程**，简称**马尔可夫链**(Markov chain)，或**马氏链**.

对马氏链，(21.1)式有如下等价形式：对任意 n 及状态 $i_0,i_1,\cdots,i_{n+1}$，只要 $P(X_j=i_j,j=0,1,\cdots,n)>0$，就有

$$P(X_{n+1}=i_{n+1}|X_0=i_0,\cdots,X_n=i_n)=P(X_{n+1}=i_{n+1}|X_n=i_n).$$

当 $T=[0,\infty)$ 而 E 可列时，X_j 为 $X(t_j)$ 即可，这里 $t_0<t_1<\cdots<t_n<t_{n+1}$(21.1)的一般化见(24.3.1).

例 21.2.4 设 $Y_1,Y_2,\cdots$ 为独立列，则 $X_n=\sum_1^n Y_i,n=1,2,\cdots$ 是马氏链. 由此知随机游动中过程 (X_n) 是一马氏链. 因若令

$$Y_i=\begin{cases}1, & \text{若第 } i \text{ 次游动向右},\\ -1, & \text{反之}.\end{cases}\qquad i=1,2,\cdots$$

则诸 Y_i 独立且 $X_n=\sum_1^n Y_i$.

3. 独立增量过程

定义 21.2.5 称 $X_T=\{X_t,t\geqslant 0\}$ 为**独立增量过程**(independent increment process)，如 $\forall$ n 及 $t_0=0\leqslant t_1<t_2<\cdots<t_n$，增量 $X_{t_2}-X_{t_1},X_{t_3}-X_{t_2},\cdots,X_{t_n}-X_{t_{n-1}}$ 相互独立.

独立增量过程的重要子类是泊松过程和维纳(Wiener)过程.

4. 鞅

定义 21.2.6 设 $T=[0,\infty)$ 或 $=\{0,1,2,\cdots\}$. 称实值过程 X_T 为 $(\mathscr{F}_t)$ **鞅**(martingale)，如

(1) $(\mathscr{F}_t)$ 是**参考族**(Peference family)，即 $\mathscr{F}$ 的子 σ 代数族，且递增、右连续：$\mathscr{F}_t\subset\mathscr{F}_s$，如 $t<s$； $\mathscr{F}_t=\mathscr{F}_{t+}\triangleq\bigcap_{s>t}\mathscr{F}_s$.

(2) X_T 为 $(\mathscr{F}_t)$ **适应**(adapted to $(\mathscr{F}_t)$)，即 $X_t\in\mathscr{F}_t$，$\forall$ $t\in T$.

(3) $E|X_t|<\infty,t\in T$.

(4) $E(X_t|\mathscr{F}_s)=X_s$，a. s. $\forall$ $s,t(>s)\in T$.

鞅始于 20 世纪 30 年代末至 50 年代初，是现代随机过程理论中最活跃和最富于成果的分支. 在马尔可夫过程、点过程、估计理论、随机系统优化与控制及决策管理等领域中有广泛应用. 鞅的收敛性与随机微积分概念及运算是研究的两个主要内容.

例 21.2.7 设 $\{Y_n,n\geqslant 0\}$ 为随机变量列 (r,v,s)，$\mathscr{F}_n=\sigma(Y_0,Y_1,\cdots,Y_n)$. X 为随机变量且 $E|X|<\infty$. 令 $X_n=E(X|\mathscr{F}_n)$，则 $\{X_n,n\geqslant 0\}$ 是 $(\mathscr{F}_n)$ 鞅. 此鞅称**杜布鞅过程**(Doob's martingale process).

例 21.2.8 设 $\{Y_n,n\geqslant 0\}$ 为独立列，$EY_n=0,n\geqslant 0$，又设 g_k 是 k 元博雷尔可测函数，$b_n=g_n(Y_0,Y_1,\cdots,Y_{n-1})$，且 $E|b_n|<\infty$，$n\geqslant 1$. 令 $X_n=x_0+\sum_1^n b_kY_k$，$x_0=\text{const}$，$\mathscr{F}_n=\sigma(Y_0,\cdots,Y_n)$，$n\geqslant 1$，则 $\{X_n,n\geqslant 1\}$ 是 $(\mathscr{F}_n)$ 鞅.

事实上由 $E|b_k|<\infty,E|Y_k|<\infty$ 及 $x_0=\text{const}$ 可知 $E|X_n|<$

∞，又 $X_{n+1}=X_n+b_{n+1}Y_{n+1}$，故由条件期望性质及 Y_{n+1} 与 $\mathscr{F}_n$ 独立，

$$\begin{aligned}E(X_{n+1}|\mathscr{F}_n) &= E(X_n|\mathscr{F}_n)+E(b_{n+1}Y_{n+1}|\mathscr{F}_n)\\&= X_n+b_{n+1}E(Y_{n+1}|\mathscr{F}_n)\\&= X_n+b_{n+1}EY_{n+1}=X_n\end{aligned}$$

21.3 随机过程的可分性

21.3.1 过程可分性的定义

可分性是为研究连续参数随机过程样本函数的性质（连续性、可微性和有界性等）而引入的概念. 由于概率空间 $(\Omega,\mathscr{F},P)$ 中 $\mathscr{F}$ 只保证对可列多个事件的并，交及求余封闭，而像 ω-集

$$B=\{\omega:\sup_{a\leqslant t\leqslant b}X_t(\omega)\leqslant c\}=\bigcap_{t\in[a,b]}\{\omega:X_t(\omega)\leqslant c\}$$

涉及非可列的无穷多个事件的并，从而就不能保证 B 是一个事件（$B\in\mathscr{F}$）.

可分性引入就解决了这类问题.

约定简记 ω-集 $\{\omega:X_t(\omega)\leqslant c\}$ 为 $(X_t\leqslant c)$，$\{\omega:X_t(\omega)\in A\}$ 为 $(X_t\in A)$ 等等.

定义 21.3.1 称函数 $x(t),t\in T$ 关于集 R 是**可分的**(separable)，如 $x_T\subset\overline{x}_R$，也即对任一 $t\in T$，可找到点列 $\{r_i\}\subset R$（允许某个 $r_i=r_j,i\neq j$），使同时有

$$r_n\to t,x(r_n)\to x(t),n\to\infty.$$

称 R 为此函数的**可分集**(separable set).

定义 21.3.2 称过程 X_T 关于 R 可分，如存在一概率为零的集 N，使对任一 $\omega\notin N$，样本函数 $X_t(\omega),t\in T$ 关于 R 可分. 此时称 R 为 X_T 的可分集，N 为 X_T 的例外集.

称过程 X_T 是可分的，如存在 T 的可列稠子集 R，使它关于 R

可分. 如 X_T 关于 T 的任一可列稠子集都是可分的，则称过程 X_T 是**完全可分的**(fully separable).

定义 21.3.3 样本函数连续(右连续、左连续)的过程称为**连续（右连续、左连续）过程**（continuous/right-continuous/left-continuous/process).

例 21.3.4 连续过程是完全可分的. 下一过程 $\{X_t, t\in[0,1]\}$ 是不可分的：

$$X_t(\omega)=\begin{cases}1, & 如\ t=\omega,\\ 0, & 如\ t\neq\omega,\ t,\omega\in[0,1].\end{cases}$$

这里概率空间为 $([0,1],\mathscr{L}[0,1],L)$，$\mathscr{L}[0,1]$ 记 $[0,1]$ 中勒贝格可测集全体，L 为 $\mathscr{L}[0,1]$ 上勒贝格测度.

如 X_T 是可分的，则 $(X_t\in A,\forall t\in T)\in\mathscr{F}$，其中 A 为任一有界闭集；$(X_t$ 在 T 有界$)\in\mathscr{F}$；$(X_t$ 在 t_0 连续)、$(|X_T-C|\leqslant\epsilon)$、$(X_t$ 在 $[a,b]$ 上均匀连续)，$[a,b]\subset T$ 及 $(\inf_{t\in T}X_t\leqslant c)$ 均为事件，$\forall c\in R_1$，从而 $\inf X_t$ 是随机变量. 对 $\sup X_t$，$\overline{\lim}X_t$ 及 $\underline{\lim}X_t$ 可建立类似结论. 可见，对连续参数过程的研究，可分性是一个最起码也是最重要的要求.

21.3.2 可分修正

并不是所有连续参数过程都是可分的. 能不能对过程作些改造，使它变成可分的，而此改造不影响过程的概率性质.

定义 21.3.5 如 $P(X_t=Y_t)=1,\forall t\in T$，称两过程 X_T 和 Y_T 是**随机等价**的（random equivalent），也称 Y_T 是 X_T 的**修正**(modification).

定理 21.3.6 杜布定理 任一过程必存在可分的随机等价过程(称为原过程的**可分修正**(separable modification).

这样，当研究只涉及过程的有限维分布函数时，总不妨设原过程是可分的.

定义 21.3.7 如$\forall s\in T$及$\varepsilon>0$,有$\lim\limits_{t\to s}P(|X_t-X_s|>\varepsilon)=0$,称$X_T$为**随机连续的**(random continuous),或**依分布连续的**(continuous with distribution).

以$t\uparrow s[\downarrow s]$易$t\to s$时,如相应的极限等式成立,称$X_T$为**左[右]随机连续的**(left(right)random continuous).

定理 21.3.8 可分过程如随机连续必完全可分.

21.4 样本函数连续性与阶梯性

为使函数在$[0,\infty)$上连续,只要它在任一$[a,b]$,$0\leqslant a<b$上连续即可.故以下设$T=[a,b]$.

定理 21.4.1 设R是X_T的可分集,N为例外集如对每$\omega\notin N$样本函数$X_t(\omega)$,$\omega\notin N$在R上一致连续,则在$T=[a,b]$上一致连续.

定理 21.4.2 设对可分过程X_T存在三个正实数a,c和ε,使$E|X_t-X_{t+\Delta}|^a\leqslant c|\Delta|^{1+\varepsilon}$,$\forall\ t,t+\Delta\in[a,b]$则$X_T$完全可分且过程存在连续修正过程,即$X_T$的几乎所有的样本函数(即以概率1)连续.

定理 21.4.3 设过程X_T增量$\Delta X_t\triangleq X_{t+\Delta}-X_t$,$t,t+\Delta\in[a,b]$是零均值正态的,且$\exists\ c\geqslant 0$,$\varepsilon>0$使$V(\Delta X_t)\leqslant c|\Delta|^{\varepsilon}$,则称$X_T$样本函数以概率1在$[a,b]$均匀连续.

定义 21.4.4 称普通实值函数$x(t)$,$a\leqslant t\leqslant b$为**阶梯的**(step),如存在$T=[a,b]$的有穷分割D:$a=t_0<t_1<\cdots<t_n=b$,使在每个(t_i,t_{i+1})中$x_t=c_i$(常数),而$x_{t_i\mp0}$存在(在t_0,t_n分别只有右、左极限)且不等,而x_{t_i}等于其中之一.上述t_i,$i=1,2,\cdots,n$称为X_T的**跳跃点**(jump point).称函数$x(t)$,$t\in[a,b]$是**几乎阶梯的**(almost step),如除去$[a,b]$上有穷多个点外,它是阶梯的.

定理 21.4.5 设X_T为可分实正态,$T=[a,b]$且$m(t)=EX_t$

连续，$|\Gamma(s,t)-\Gamma(s,s)|\leqslant c|t-s|^{\epsilon}$，其中 $c\geqslant 0,\epsilon>0$，则 X_T 样本函数以概率 1 均匀连续.

定义 21.4.6 称函数 $x_t,a\leqslant t<\infty$ 为阶梯的[几乎阶梯的]，如存在 $b_n\uparrow\infty$，使在每个 $[a,b_n](b_n>a)$ 中为阶梯的[几乎阶梯的].

定义 21.4.7 设 $T=[a,\infty)$ 或 $[a,b]$. 如存在概率为零的集 N，使对每 $\omega\notin N$，样本函数 $X_t(\omega),t\in T$ 为阶梯的[几乎阶梯的]. 称过程 X_T 的**样本函数以概率 1 为阶梯的**(step function with probability 1)[几乎阶梯的(almost step with probability 1)].

定理 21.4.8 设 X_T 可分，$T=[a,b]$，且 $\exists\ c\geqslant 0$ 使 $P(X_s\neq X_t)\leqslant c|t-s|,\forall\ t,s\in[a,b]$，则 X_T 的样本函数以概率 1 为阶梯的.

定理 21.4.9 设 X_T 可分，$T=[a,b]$，且 $\exists\ c\geqslant 0$ 使 T 中任意有限点列 $t_0<t_1<\cdots<t_n$ 都有 $\sum_1^n P\left(X_{t_{i-1}}\neq X_{t_i}\right)\leqslant c$. 则此过程的样本函数以概率 1 是几乎阶梯的. 如又设 X_T 完全可分，则样本函数以概率 1 是阶梯的.

21.5 过程的可测性

在过程的定义中，要求对固定的 $t\in T$，$X_t(\omega)$ 是 $\mathscr{F}$ 可测的(即是随机变量)，却对固定 $\omega\in\Omega$，其样本函数 $X(\cdot,\omega),\cdot\in T$ 尚没有规定可测性. 为考虑对样本函数对 t 的积分，引入过程可测性.

设 $T=[a,b]$ 或 $[0,\infty)$，$\mathscr{B}_T$ 为 T 中全体 Borel 集所成的 σ 代数，L 表 $\mathscr{B}_T$ 上的 Lebesque 测度，而 $\overline{\mathscr{B}_T\times\mathscr{F}}$ 表直积 σ 代数 $\mathscr{B}_T\times\mathscr{F}$ 关于测度 $\mu=L\times P$ 的完备化.

定义 21.5.1 称 $(\Omega,\mathscr{F})$ 上过程 X_T 为**可测的**(measurable)如

$$\{(t,\omega):X_t(\omega)\leqslant c,t\in T,\omega\in\Omega\}\in\overline{\mathscr{B}_T\times\mathscr{F}},\forall\ c\in R_1;$$

称 X_T 为博雷尔可测的，如上式右方可写为 $\mathscr{B}_T\times\mathscr{F}$.

定理 21.5.2

(1) 随机连续过程必有完全可分且可测的修正.

(2) 博雷尔可测过程的所有矩函数都是博雷尔函数.

(3) 至多有可列个间断点的过程，必有博雷尔可测修正.

在马尔可夫过程论中尚需要更强的可测性.

定义 21.5.3 如对任意 $B\in\mathscr{B}_T$，$[u,v]\subset T$，有 $\{(t,\omega):u\leqslant t\leqslant v,\omega\in\Omega,X_t(\omega)\in B\}\in\mathscr{B}_{[u,v]}\times\mathscr{N}_v^u$. 称过程 X_T 为**强可测的**(strong measurable). 其中 $\mathscr{N}_v^u=\sigma\{X_t(\omega),u\leqslant t\leqslant v\}$，可理解为在时段 $[u,v]$ 内，过程运动中所发生的事件所生成的 σ 代数.

定理 21.5.4 右连续过程是博雷尔可测的，且是强可测的.

法国 Strasbourg 学派在随机过程一般理论中把过程看成二元函数，由此区分一些特殊的二元可测类. 它们在随机积分理论中起重要作用.

定义 21.5.5 设 $(\mathscr{F}_t)$ 为 $(\Omega,\mathscr{F})$ 上参考族，$T=[0,\infty)$. 如 $\forall t>0$，$X(s,\omega)$，$0\leqslant s\leqslant t$，$\omega\in\Omega$ 是 $\mathscr{F}_t\times\mathscr{B}_{[0,t]}$ 可测. 则称 X 为 $(\mathscr{F}_t)$ **循序过程**(progressive process). 又记 $\mathscr{P}_g=\{A:\in\mathscr{B}_{[0,\infty)}\times\mathscr{F},I_A$ 为 $(\mathscr{F}_t)$ 循序过程$\}$.

$\mathscr{O}=\sigma$(右连续且有左连续的 $(\mathscr{F}_t)$ 适应过程).

$\mathscr{O}^*=\sigma$(右连续 $(\mathscr{F}_t)$ 适应过程).

$\mathscr{P}=\sigma$(左连续 $(\mathscr{F}_t)$ 适应过程).

$\mathscr{P}^*=\sigma$(左连续且有右极限的 $(\mathscr{F}_t)$ 适应过程).

$\mathscr{P}^{**}=\sigma$(连续 $(\mathscr{F}_t)$ 适应过程).

称 $\mathscr{P}_g$ $[\mathscr{O},\mathscr{P}]$ 为 $(\mathscr{F}_t)$ **循序**[**可选**(optional)，**可料**(predictable)]σ **代数**. 称 X_T 为 $(\mathscr{F}_t)$ **可选**[**可料**]**过程**，如 X_T 为 $\mathscr{O}[\mathscr{P}]$ 可测.

它们之间有下列关系：

$\mathscr{P}\subset\mathscr{O}\subset\mathscr{O}^*\subset\mathscr{P}_g$，故可料必可选，可选必循序；$\mathscr{P}=\mathscr{P}^*=$

$\mathscr{P}^{**}$;

设 $(\mathscr{F}_t)$ 为参考族，X_T 为 $(\mathscr{F}_t)$ 适应的博雷尔可测过程，则必存在一个 $(\mathscr{F}_t)$ 可选修正.

21.6 马尔可夫过程中的算子理论

马尔可夫过程的研究有两种类型问题. 一是用泛函分析中算子半群理论来研究过程的一个重要特征，转移概率；另一是研究过程本身，例如它的样本函数. 后者主要工作在 Ω 上进行，主要工具是测度论，将在第 24 章中介绍.

21.6.1 转移概率算子半群

设 $X=(X_t)_{t\geqslant 0}$ 为一齐次马尔可夫过程，$\mathscr{B}$ 是其相空间 E 上的一个 σ 代数. $P(t,x,\Gamma)$ 为 X 的转移概率，即 $P(t,x,\Gamma)=P(X_t\in\Gamma|X_0=x)\triangleq P_x(X_t\in\Gamma)$，$\Gamma\in\mathscr{B}$，记 $B(E)$ 为 E 上一切有界 $\mathscr{B}$ 一可测函数的巴拿赫空间. 令 $\|f\|=\sup\limits_{x\in E}|f(x)|$ 为 $f\in B(E)$ 的范数. 并在 $B(E)$ 上定义算子 T_t：

$$T_tf(x)=\int_E f(y)P(t,x,\mathrm{d}y),\quad x\in E,\quad f\in B(E),$$

算子性质 21.6.1 $T_t,t\geqslant 0$ 的性质

(1) 对任意 $t\geqslant 0$，T_t 是 $B(E)$ 到 $B(E)$ 的有界线性算子，且 $\|T_t\|\leqslant 1$；

(2) 对一切 $s,t\geqslant 0$，有 $T_{t+s}=T_tT_s$；

(3) 如对一切 $x\in E$ 有 $f(x)\geqslant 0$，则对一切 $x\in E$ 和 $t\geqslant 0$ 有 $T_tf(x)\geqslant 0$；

(4) 如 $f(x_0)=0$，则 $T_0f(x_0)=0$；

(5) 如对一切 $x\in E$，$\lim\limits_{n\to\infty}f_n(x)=f(x)$，$f_n\in B(E)$，且 $\sup\limits_n\|f_n\|<\infty$，则 $\lim\limits_{n\to\infty}T_tf_n(x)=T_tf(x)$.

定义 21.6.2 称满足(21.6.1(1))和(21.6.1(2))的算子簇$(T_t)_{t\geqslant 0}$为**压缩算子半群**(contraction semigroup of operators)

(21.6.1(2))的证明用到齐次转移概率科尔莫戈罗夫-查普曼(Kolmogorov-Chapman)方程.

$$P(t+s,x,\Gamma)=\int_E P(t,y,\Gamma)P(s,x,\mathrm{d}y).$$

作用于$B(E)$并满足性质 21.6.1 的任何压缩算子半群,都产生齐次转移概率,且$P(t,x,\Gamma)=T_tI_\Gamma(x)$,这里$I_\Gamma(\cdot)$为集$\Gamma$的示性函数.这样对齐次马氏过程的研究可利用半群理论.

21.6.2 无穷小算子

定义 21.6.3 设$(E,\mathscr{B})$为某可测空间,$(T_t)_{t\geqslant 0}$为作用于$B(E)$中的压缩算子半群.由公式$Af=g$决定的算子A称为半群T_t的**无穷小算子**(infinitesimal operator),若

$$\lim_{t\downarrow 0}\left\|g-\frac{T_tf-f}{t}\right\|=0,$$

A的定义域为$D_A=\{f:\in B(E)$,且对$x\in E$一致存在$\lim\limits_{t\downarrow 0}\frac{1}{t}(T_tf(x)-f(x))\in B(E)\}$

性质 21.6.4 无穷小算子性质:

(1) 集D_A依范数收敛(称强收敛)意义下的闭包与集$B_0(E)\triangleq\{f:\in B(E)$且$\lim\limits_{t\downarrow 0}\|T_tf-f\|=0\}$重合;此处$\|g\|=\sup|g(x)|$.

(2) 如$f\in D_A$,则$Af\in B_0(E)$且$T_tf-f=\int_0^t T_sAf\mathrm{d}s$;

(3) 如$f\in D_A$,则T_tf对$t\geqslant 0$强可微,且$\frac{\mathrm{d}T_tf}{dt}=AT_tf=T_tAf$;这里强可微是指强收敛意义下可微.

(4) 算子A封闭.

21.6.3 用算子 A 定义扩散过程

定义 21.6.5 设 $p(t,x,\Gamma)$为相空间$(E,\mathscr{B})$中的齐次转移概率,由性质 21.6.1 产生作用于 $B(E)$的算子半群$(T_t)_{t\geqslant 0}$,此半群由定义 21.6.3 定义无穷小算子. 称此 A 为转移概率 $\boldsymbol{p(t,x,\Gamma)}$**的无穷小算子**.

由定义 21.6.3 知

$$Af(x)=\lim_{t\downarrow 0}\frac{1}{t}\left[\int_E f(y)p(t,x,\mathrm{d}y)-f(x)\right],$$

且当此极限关于 $x\in E$ 一致存在时,$f\in D_A$.

定义 21.6.6 如对每 $x\in E$ 及 x 的任一邻域 U 有$\lim\limits_{t\downarrow 0}p(t,x,U)=1$,则称转移概率为**随机连续**.

拓扑相空间中任一随机连续转移概率都唯一决定于自己的无穷小算子.

定义 21.6.7 设 E 为拓扑空间,$\mathscr{B}$ 为其博雷尔集的 σ 代数. 如对一切 $t>0$ 和 $f\in C(E)\triangleq\{f:E$ 上有界连续实函数$\}$,成立

$$T_tf(x)=\int_E p(t,x,\mathrm{d}y)f(y)\in C(E),$$

则称齐次 $p(t,x,\Gamma),\Gamma\in\mathscr{B}$ 为**费勒转移概率**. 有此转移概率的马尔可夫过程称为**费勒过程**(Feller process).

定理 21.6.8 如满足第一可数公理,则随机连续的费勒转移概率的 $C(E)$上的无穷小算子唯一决定转移概率.

性质 21.6.9 对紧致集 E 上的右连续费勒过程,有如下结论:

(1) 它必是强马尔可夫过程(见定义 24.3.5);

(2) 设 A 是其无穷小算子,$f\in D_A$,则对一切 $t\geqslant 0,x\in E$ 有

$$E_xf(X_t)-f(x)=E_x\int_0^t Af(X_s)\mathrm{d}s;$$

(3) 对任一**马尔可夫**时间(见定义 24.3.4)τ,当 $E_x\tau<\infty$ 时,有

$$E_x f(X_\tau) - f(x) = E_x\int_0^\tau Af(X_s)\mathrm{d}s;$$

邓肯公式(Дынкин formula)

定义 21.6.10 对固定 $x\in E$,令 τ_n 是 x 的邻域 U_n 的首中时(hitting time),且 $U_n\downarrow\{x\}$.

$D_n^x\triangleq\{f:\in C(E)$,且使 $\mathscr{U}f(x)\triangleq\lim\limits_{n\to\infty}[E_xf(X_{t_n}-f(x)]/E_x\tau_n$ 存在$\}$

$D_{\mathscr{U}}\triangleq\bigcap\limits_{x\in E}D_n^x$,则当 $f\in D_{\mathscr{U}}$ 时对一切 $x\in E$,$\mathscr{U}f(x)$ 有定义.称算子 $\mathscr{U}$ 为过程 X 的**特征算子**(characteristic operator).

定理 21.6.11 半紧致(semi-compact)E 上右连续 Feller 过程的特征算子是它在 $C(E)$ 上无穷小算子的扩张:$D_A\subset D_{\mathscr{U}}$,且当 $f\in D_A$,$Af(x)=\mathscr{U}f(x)$,$x\in E$.

性质 21.6.12 连续费勒过程的特征算子的性质

(1) 局部性:在 x_0 某邻域内 $f_1(x)=f_2(x)$,则 $\mathscr{U}f_1(x)=\mathscr{U}f_2(x)$,即 $\mathscr{U}f(x_0)$ 的值只依赖 $f(x)$ 在 x_0 任一邻域中的值;

(2) 线性;

(3) 非负定性:如 $f(x)$ 在 x_0 达到相对极小(即对 x_0 的某邻域 U_{x_0},$f(x)\geqslant f(x_0)$,$x\in U_{x_0}$),则 $\mathscr{U}f(x_0)\geqslant 0$;

(4) 椭圆性:$\mathscr{U}$ 是广义二阶椭圆型微分算子的一般化.具体说,若 $\mathscr{U}$ 在 U_{x_0} 内对连续函数 φ 及 φ^2 有定义,f 在 y_0 邻域连续二次可微,$y_0=\varphi(x_0)$,则在点 x_0 上 $\mathscr{U}$ 对 $f(\varphi)$ 有定义,且

$$\mathscr{U}f(\varphi)=a\frac{\mathrm{d}f}{\mathrm{d}\varphi}+\frac{1}{2}b\frac{\mathrm{d}^2f}{\mathrm{d}\varphi^2},\quad b>0$$

其中 $a=\mathscr{U}(\varphi-\varphi^{(0)})$,$b=\mathscr{U}(\varphi-\varphi^{(0)})^2$,$\varphi^{(0)}=\varphi(x_0)$,特别可取 $\varphi(x)=x$.

22　二阶矩过程

22.1　定义与例

22.1.1　引言与定义

二阶矩过程是一类重要的随机过程，在物理学、生物学、通讯与控程、系统工程与管理科学等方面，都有广泛的应用.

随机过程的概率性质由它的分布函数族完全决定，但在实际问题里，要求出它们是十分困难的. 由统计方法得到过程的一阶和二阶矩，还是可能的. 对那些二阶矩存在的过程来说，通过二阶矩（主要是协方差函数）的研究可以对该过程的特性获得了解. 在不少实际问题里，这些了解就足够了. 更何况要对某些过程，如正态过程，一、二阶矩就能完全确定有限维分布.

最常遇到的二阶矩过程，有正态过程、（宽）平稳过程、一大类独立增量过程，以及正交增量过程.

定义 22.1.1　如过程 $X_T=\{X_t,t\in T\}$ 对每个 $t\in T$，X_t 的方差都存在，则称它是**二阶矩过程**.

由矩的性质：高阶矩存在则低阶矩一定存在及柯西-施瓦兹不等式可知，二阶矩过程的均值函数与协方差函数都存在.

如 X_T 是二阶矩过程，令 $\widetilde{X}_t=X_t-m(t)$，$t\in T$，则 $\widetilde{X}_T=\{\widetilde{X}_t,t\in T\}$ 也是二阶矩过程，且其均值函数恒为 0. 因此，今后一般总设二阶矩过程为零均值的.

22.1.2　例

例 22.1.2　余弦波过程(cosine wave process)　下一过程称

为余弦波过程，它是一个二阶矩过程.

$$X_t = A\cos(\omega t + \theta), \quad t \geqslant 0,$$

其中振幅 A 与角频率 ω 取常数，相位 θ 是一个在 $(-\pi,\pi)$ 内的均匀分布的随机变量.

解 $E|X_t|^2 = \int_{-\pi}^{\pi} A^2\cos^2(\omega t + \alpha)\cdot\frac{1}{2\pi}\mathrm{d}\alpha \leqslant A^2, \forall\, t \geqslant 0$，故 X_t 为二阶矩过程.

$EX_t = \int_{-\pi}^{\pi} A\cos(\omega t + \alpha)\,\frac{1}{2\pi}\mathrm{d}\alpha = 0$，从而

$$\Gamma(s,t) = EX_sX_t = \int_{-\pi}^{\pi} A^2\cos(\omega s + \alpha)\cos(\omega t + \alpha)\cdot\frac{1}{2\pi}\mathrm{d}\alpha$$

$$= \frac{1}{2}A^2\cos\omega(s - t)$$

它只与时间差 $s-t$ 有关，此过程也是平稳过程.

例 22.1.3 设 $X_t = X_0 + Vt, a \leqslant t \leqslant b$，其中 X_0 与 V 独立且同分布 $N(0,1)$，则此过程是二阶矩过程. 事实上由正态分布性质：(X,Y) 是二元正态分布的充要条件为对任一非平凡的线性组合 $\alpha X + \beta Y$ 是一元正态的. 知 X_t 是正态变量，故方差存在，从而是二阶矩过程，易得 $m(t) \equiv 0, \Gamma(s,t) = EX_sX_t = 1 + st$，它不是平稳过程.

22.2 二阶矩过程的性质

22.2.1 协方差函数性质

性质 22.2.1 二阶矩过程协方差函数性质

(1) 复对称性或埃尔米特(Hermit)性：$\Gamma(s,t) = \overline{\Gamma(t,s)}, s,t \in T$. 如过程是实的，此即普通的对称性.

(2) 非负定性：对任意有限个 $t_1, t_2, \cdots, t_n \in T$ 和 T 上的普通复值函数 θ，有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k,t_j)\theta(t_k)\overline{\theta(t_j)}\geqslant 0.$$

注 22.2.2 由非负定性可推出埃尔米特性.可见非负定性是协方差函数的本质特性.

定理 22.2.3 已知 $T\times T$ 上非负定函数 $\Gamma(s,t)$,则一定存在二阶矩过程 $X_T=\{X_t,t\in T\}$(事实上还可以是正态过程),使其协方差函数就是 $\Gamma(s,t)$.若 $\Gamma(s,t)$是 $T\times T$ 上实函数,则相应的 X_T 也是实的.

定理 22.2.4 已知 T 上函数 $m(t)$及 $T\times T$ 上非负函数 $\Gamma(s,t)$,则存在唯一的有限维分布族,使其为正态过程的有限维分布族.

22.2.2 二阶矩随机变量空间

1. H 空间的构成与赋范

定义 22.2.5 同一概率空间$(\Omega,\mathscr{F},P)$上的具有穷二阶矩的随机变量所组成的空间,称为**二阶矩随机变量空间**,简称**二阶矩空间**(second order space),记为 H.

定理 22.2.6 H 是线性空间;如 $X_i\in H$,$c_i=$const.(可为复值),$i=\overline{1,n}$,则 $\sum\limits^{n}c_iX_i\in H$.

定理 22.2.7 对任意 $X,Y\in H$,令$(X,Y)=EXY$,则(X,Y)为 H 中的内积:即如 $X_i,Y_i,X,Y\in H$,$i=1,2$,则

(1) $(X,X)\geqslant 0$,$X=0$(a.e)$\Leftrightarrow(X,X)=0$;

(2) 复对称性:$(Y,X)=\overline{(X,Y)}$;

(3) 复线性:$(c_1X_1+c_2X_2,Y)=c_1(X_1,Y)+c_2(X_2,Y)$

$$(X,c_1Y_1+c_2Y_2)=\bar{c}_1(X_1,Y_1)+\bar{c}_2(X,Y_2).$$

定理 22.2.8 对每 $X\in H$ 令 $\|X\|^2=(X,X)(=E|X|^2)$,则 $\|X\|$ 是 H 中范数(norm):即如 $X_1,X_2,X\in H$,则

(1) $\|X\|\geqslant 0$;$X=0$(a.e.)$\Leftrightarrow\|X\|=0$;

(2) $\|cX\| = |c| \|X\|$;

(3) $\|X_1+X_2\| \leqslant \|X_1\| + \|X_2\|$.

综上可知,H 为线性赋范空间和有内积的线性空间.

2. 空间 H 的均方收敛与完备性

定理 22.2.9 对任 $X,Y\in H$,令 $d(X,Y)=\|X-Y\|$,则 $d(X,Y)$是 H 中的距离,即对 $X,Y,Z\in H$,

(1) 非负与零性质:$d(X,Y)\geqslant 0$;$X=Y$(a. e.)$\Leftrightarrow d(X,Y)=0$.

(2) 对称性:$d(X,Y)=d(Y,X)$.

(3) 三角不等式:$d(X,Z)\leqslant d(X,Y)+d(Y,Z)$;

定义 22.2.10 设 $X_n,X\in H$,如 $d(X_n,X)\to 0(n\to\infty)$,则称 X_n **均方收敛**(mean squane convergence)于 X,记 $\text{l.i.m}X_n\to X$;,如 $\{X_n\}_1^{\infty}\subset H$,$d(X_n,X_m)\to 0(n,m\to\infty)$,则称$\{X_n\}$为均方基本列或均方柯西列.

3. 均方收敛的性质

均方收敛必为依概率收敛,但反之不真. 但如$|X_n|<Y\in H$,则由 X_n 依概率收敛可推出均方收敛.

性质 22.2.11 均方收敛性的几个基本结论:

设 $Y_n,Y,X_n,X\in H,n\geqslant 1$,且 $\text{l.i.m}X_n=X$,$\text{l.i.m}Y_n=Y$,则

(1) $\lim EX_n=EX=E(\text{l.i.m}X_n)$.

(2) $\lim E|X_n|^2=E|X|^2$.

(3) $\lim\limits_{n,m\to\infty}(X_n,Y_m)=(X,Y)$,即 $\lim\limits_{n,m\to\infty}EX_nY_m=EXY$.

(4) $\text{l.i.m}(aX_n+bY_n)=aX+bY$,$\forall\ a,b\in R$.

(5) 均方极限唯一:如又有 $\text{l.i.m}X_n=Y$,则 $X=Y$(a. e.).

定理 22.2.12(均方收敛的柯西准则(Cauchy criterion for mean square convergence) $\{X_n\}(\subset H)$均方收敛的充要条件是 $\lim\limits_{n,m\to\infty}E|X_n-X_m|^2=0$.

定理 22.2.13 完备性 设$\{X_n\}$为 H 中均方基本列,则$\exists\ X$

$\in H$,使 $\text{l.i.m}X_n = X$. 从而 H 为巴拿赫空间(完备的线性赋范空间);也是希尔伯特空间(完备的内积空间).

定理 22.2.14 均方收敛准则或洛也夫准则(Loéve criterion) 设 $X_n, X \in H$,则 $\text{l.i.m}X_n = X$ 的充要条件是 $\lim\limits_{n,m\to\infty} E(X_n, \overline{X}_m) = \lim\limits_{n,m}(X_n, X_m) = c$(常数),此时有 $c = \|X\|^2$.

定理 22.2.15 设 $X_n, X \in H$, $\text{l.i.m}X_n = X$, $f(u)$是一确定的函数,设 $f(X_n), f(X) \in H$,且满足利普希茨(Lipschitz)条件:存在常数 M,使 $|f(u) - f(v)| \leqslant M|u - v|$. 则 $\text{l.i.m}f(X_n) = f(X)$.

推论 22.2.16 若 $f(u)$和 $f'(u)$都存在且有界, $\text{l.i.m}X_n = X$,则 $\text{l.i.m}f(X_n) = f(X)$.

若 $\text{l.i.m}X_n = X$,则对任意有限的 t,有 $\text{l.i.m}\exp\{itX_n\} = \exp\{itX\}$,从而 X_n 的特征函数收敛于 X 的特征函数;相应有分布函数的收敛性.

定理 22.2.17 均方收敛的大数定理 设$\{X_n\}$为 H 中独立同分布的点列(随机变量列),则 $\text{l.i.m}\,\frac{1}{n}\sum^{n} X_i = a = EX_1$.

22.2.3 过程的随机分析性质

二阶矩过程 $X_T = \{X_t, t \in T\}$,对每 $t \in T$, $X_t \in H$,故 X_T 为空间 H 中一个曲线(如 T 连续)或一个点列(如 T 离散). 本段对 $T \subset R$,且连续的二阶矩过程,进一步讨论其均方收敛意义下的连续性,可微性与可积性,并用均方差函数的表现给出判别准则.

定义 22.2.18 称 $X_T \subset H$ 在 $t_0 \in T$ **均方连续的**(mean square continuous),如 $\text{l.i.m}\limits_{t\to t_0} X_t = X_{t_0}$;如对$\forall\ t \in T$, X_t 在 t 均方连续,则称 X_t 为**均方连续**.

定理 22.2.19 均方连续准则(mean square continuity criterion) $X_T(\subset H)$在 $t_0 \in T$ 均方连续$\Leftrightarrow X_T$ 的协方差函数

$\Gamma(s,t)$在(t_0,t_0)连续.

又,$\Gamma(s,t)$在$T\times T$连续的充要条件是它在(t,t),$t\in T$连续.

定义 22.2.20 称$X_T\subset H$在$t_0\in T$**均方可微**(mean square differentiable),如$\underset{h\to 0}{\text{l.i.m}}\ \frac{1}{h}(X_{t_0+h}-X_{t_0})\triangleq X_{t_0}'$存在.如$X_T$在每$t\in T$为均方可微,则称$X_T$一次均方可微,循此可定义二次均方可微,等等.

定理 22.2.21 均方可微准则(mean square differentiability criterion) $X_T(\subset H)$在$t_0(\in T)$均方可微$\Leftrightarrow X_T$的协方差函数$\Gamma(s,t)$在(t_0,t_0)广义二次可微,即下面的均方极限存在:

$$\lim_{h,h'\to 0}\frac{1}{hh'}[\Gamma(t_0+h,t_0+h') - \Gamma(t_0+h,t_0)-\Gamma(t_0,t_0+h')+\Gamma(t_0,t_0)] \quad (22.1)$$

定理 22.2.22 如$\Gamma(s,t)$的一阶偏导存在,$\frac{\partial^2\Gamma}{\partial s\partial t}$存在且连续,则$\Gamma(s,t)$在$(s,t)$广义二次可微,即以$s+h$和$t+h'$分别换(22.1)式中的$t_0+h$和$t_0+h'$,所得的均方极限存在.反之则$\frac{\partial\Gamma}{\partial s}$,$\frac{\partial\Gamma}{\partial t}$,$\frac{\partial^2\Gamma}{\partial s\partial t}$和$\frac{\partial^2\Gamma}{\partial t\partial s}$存在,且$EX_s'X_t=\frac{\partial\Gamma(s,t)}{\partial s}$,$EX_sX_t'=\frac{\partial\Gamma(s,t)}{\partial t}$,$EX_s'X_t'=\frac{\partial^2\Gamma(s,t)}{\partial s\partial t}=\frac{\partial^2\Gamma(s,t)}{\partial t\partial s}$.

定义 22.2.23 均方可积(mean square integrable) 设$X_t\in H$,$f(t)$为$T=[a,b]$上普通函数,T的一组分点:$a=t_0<t_1<\cdots<t_n=b$,$\Delta_n=\max\limits_k(t_k-t_{k-1})$,作和

$$Y_n=\sum_{k=1}^{n}f(u_k)X_{u_k}(t_k-t_{k-1}),\quad u_k\in[t_{k-1},t_k]$$

由定理22.2.6,$Y_n\in H$,如$\Delta_n\to 0$时Y_n均方收敛,且其极限与T的分割无关,则称$f(t)X_t$在$[a,b]$上黎曼**均方可积**.极限记为

$$(ms)\int_a^b f(t)X_t\mathrm{d}t.$$

$f(t)X_t$ 在 $T=[a,\infty)$ 上(Riemann)均方积分 $\int_a^\infty f(t)X_t\mathrm{d}t$ 定义为 $\underset{b\to\infty}{\mathrm{l.i.m}}\int_a^b f(t)X_t\mathrm{d}t$,如 $f(t)X_t$ 在 $[a,b]$ 均方可积,且其积分的均方极限存在.

定理 22.2.24 均方可积准则(m.s. integrability criterion) 设 $X_T\subset H$,$\Gamma(s,t)$ 为其协方差函数,则 $f(t)X_t$ 在 $[a,b]$ 均方可积 $\Leftrightarrow$ 普通积分 $\int_a^b\int_a^b f(s)\overline{f(t)}\Gamma(s,t)\mathrm{d}s\mathrm{d}t$ 存在.此时

$$E\int_a^b f(t)X_t\mathrm{d}t=\int_a^b f(t)EX_t\mathrm{d}t,$$

$$E\left|\int_a^b f(t)X_t\mathrm{d}t\right|^2=\int_a^b\int_a^b f(s)\overline{f(t)}\Gamma(s,t)\mathrm{d}s\mathrm{d}t.$$

可证相应于分析中的结论(连续、可微、可积间的关系与性质)均成立.例如当 X_t 在 $[a,b]$ 均方连续,则 $Y_t\triangleq(ms)\int_a^t X_s\mathrm{d}s$ 在 $[a,b]$ 均方连续、均方可微,且 $Y_t'=X_t$.

应用 22.2.25 直线上的布朗运动(Brown motion on a line) 设直线上某点受其它质点碰撞而运动,其到时刻 t 所受的碰撞次数 $N(t)$ 为一强度为 μ 的泊松过程.又设质点每受一次碰撞将只改变运动方向,即速度反向,为常数 $\pm v$,以 V_t 表示 t 时速度,则

$$V_t=V_0(-1)^{N(t)},\quad V_0\sim\begin{bmatrix}-v & v\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}.$$

又 $\Gamma_V(s,t)=R_V(s,t)=V^2\mathrm{e}^{-2\mu|t-s|}$,它是零均值平稳过程.以 X_t 表示此质点到时刻 t 在直线上的总位移,则 $X_t=\int_0^t V_s\mathrm{d}s,t\geqslant 0$.由定理 22.2.24 有

$$E|X_t|^2=\int_0^t\int_0^t EV(t_1)V(t_2)\mathrm{d}t_1\mathrm{d}t_2$$

$$=2\int_0^t\mathrm{d}t_1\int_0^{t_1}\mathrm{d}t_2\{v^2\mathrm{e}^{-2\mu(t_1-t_2)}\}$$

$$= \frac{v^2}{2\mu^2}(e^{-2\mu t} - 1 + 2\mu t).$$

特别，$\lim\limits_{t\to\infty} E|X_t|^2/t = \frac{v^2}{2\mu^2}$，$\lim\limits_{t\to 0} \frac{1}{t^2}E|X_t|^2 = v^2$.

22.3 几类重要的二阶矩过程

正态过程、(宽)平稳过程、二阶矩独立增量过程及正交增量过程，是二阶矩过程最重要的子类.

22.3.1 正态过程

定义 22.3.1 设 X_T 的任一有限维分布函数都是多维正态分布，则称 X_T 为**正态过程**(normal process)或**高斯过程**.

例 22.3.2 设 $X_T = \zeta\cos\theta t + \eta\sin\theta t$，$0 \leqslant t \leqslant 1$，其中 ζ、η 独立且同 $N(0, \sigma^2)$ 分布，θ 为常数. 由定义 21.2.2 后所指出的正态分布的性质：$(X_{t_1}, X_{t_2}, \cdots X_{t_n})$ 为 n 维正态分布的充要条件是对任意不全为 0 的实数 $a_1, \cdots, a_n$，和 $\sum^n a_i X_{t_i} = \left(\sum^n a_i\cos\theta t_i\right)\zeta + \left(\sum^n a_i\sin\theta t_i\right)\eta$ 是一元正态的. 由于 ζ 与 η 独立且正态，故 (ζ, η) 为二元正态，从而 ζ 与 η 的线性和是一元正态的，即此条件满足.

易求得此过程(有随机振幅的随机振动过程)的均值函数为 0，而

$$\Gamma(s, t) = \sigma^2\cos\theta(s - t).$$

它只与 $s-t$ 有关，由定义 22.3.3 知其还是平稳过程.

正态过程的进一步讨论见(22.4).

22.3.2 (宽)平稳过程

定义 22.3.3 设 X_T 为二阶矩过程，且 $m(t) = m$(常数)，

$\Gamma(s,t)$只与$s-t$有关而与s,t本身值无关，则称X_T为（宽）**平稳过程**(stationary process in wide sense)，也称弱平衡过程简称**平稳过程**.

例 22.3.2 是（宽）平稳过程的例子．再看下例．

例 22.3.4 随机电报信号(random telegraph signal) 设$T=[0,\infty)$，过程X_T满足

(1) 任一时刻$t(\geqslant 0)$，X_t取值-1或1，且取此两值为等可能，即$X_t\sim\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$；

(2) 在$(0,t]$时段内X_T变值的次数N_t，$t\geqslant 0$，有参数为λt的泊松分布，即

$$p(N_t=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad \lambda>0,\ t>0,\ k=0,1,2,\cdots;$$

(3) 初值X_0与N_t独立．

则称X_t为随机电报信号．

可写$X_t=X_0(-1)^{N_t}$．(X_{t_1},X_{t_2})的特征函数$E\exp(i(\mu_1X_{t_1}+\mu_2X_{t_2}))=e^{i(\mu_1+\mu_2)}P_{11}+e^{-i(\mu_1+\mu_2)}P_{-1-1}+e^{i(\mu_1-\mu_2)}P_{1-1}+e^{-i(\mu_1-\mu_2)}P_{-11}$，其中$P_{ij}=P(X_{t_1}=i,X_{t_2}=j)$．易算得（设$t_2>t_1$）$P_{11}=P(X_{t_1}=1,N_{t_2}-N_{t_1}=\text{偶数})=\frac{1}{2}P(N_{t_2}-N_{t_1}=\text{偶数})$．而$P_{-1-1}=\frac{1}{2}\alpha$，$P_{1-1}=P_{-11}=\frac{\beta}{2}$，其中$\alpha=P(N_{t_2}-N_{t_1}=\text{偶数})$，$\beta=1-\alpha$．于是前述特征函数等于

$$\cos(\mu_1+\mu_2)\alpha+\cos(\mu_1-\mu_2)\beta=\cos\mu_1\cos\mu_2-(\alpha-\beta)\sin\mu_1\sin\mu_2$$

$$=\cos\mu_1\cos\mu_2-\sin\mu_1\sin\mu_2\sum_{n=0}^{\infty}(-1)^nP(N_{t_2}-N_{t_1}=n)$$

$$=\cos\mu_1\cos\mu_2-\sin\mu_1\sin\mu_2\exp(-2\lambda(t_2-t_1)).$$

由特征函数与矩的关系及$EX_{t_1}=EX_{t_2}=0$，知$\Gamma(t_1,t_2)=EX_{t_1}$

$X_{t_2}=e^{2\lambda(t_2-t_1)}$，$t_2>t_1$. 协方差存在，且知此过程是(宽)平稳的.

平稳过程的进一步讨论见第 23 章.

22.3.3 二阶矩独立增量过程

定义 22.3.5 设 X_t 为随机过程，$T=[0,\infty)$. 如对任意有限个 $0\leqslant t_1<t_2\leqslant\cdots<t_n$，增量 $X_{t_2}-X_{t_1}, X_{t_3}-X_{t_2},\cdots,X_{t_n}-X_{t_{n-1}}$ 相互独立，则称 X_T 为**独立增量过程**(independent increment process).

独立增量过程并没有规定二阶矩存在. 二阶矩存在的独立增量过程是用得最多也最重要的子类，像泊松过程、布朗运动过程(详见第 27 章和第 26 章).

如定义 22.3.5 中规定过程有零初值：$X_0=0$ 或规定 $X_{t_1}, X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}}$ 相互独立，则利用例 21.2.4 知它一定是马尔可夫过程(详见第 24 章).

对 $T(\subset R_1)$ 离散时，可仿定义 22.3.5 定义独立增量过程.

22.3.4 正交增量过程

定义 22.3.6 设 $T=[0,\infty)$，X_T 为二阶矩过程，且对 $\forall\ 0\leqslant t_1<t_2\leqslant t_3<t_4$，$E(X_{t_2}-X_{t_1})\overline{(X_{t_4}-X_{t_3})}=0$，则称 X_t 为**正交增量过程**(orthogonal increment process).

如 $T=[a,b]$，且规定 $X_a=0$，则取 $t_1=a, t_2=t_3=s, t_4=t$ 时正交增量过程 X_T 有

$$EX_s\overline{(X_t-X_s)}=0,\quad s<t.$$

从而 $EX_s\overline{X}_t=E|X_s|^2\triangleq F(s)$. 此时 $\Gamma(s,t)=F(s)$.
一般地 $\Gamma(s,t)=F(s\wedge t)$，其中 $s\wedge t\triangleq\min\{s,t\}$.

可见它不是宽平稳过程，当 $s<t$ 时

$$\begin{aligned}E|X_t-X_s|^2&=\Gamma(t,t)-\Gamma(t,s)-\Gamma(s,t)+\Gamma(s,s)\\&=F(t)-F(s),\end{aligned}$$

故 $F(t)$是单调不减函数.

零均值、二阶矩独立增量过程,是正交增量过程.

22.4 正态过程及其应用

高斯过程或正态过程 X_T 的均值数和协方差函数 $\Gamma(s,t)=E\{X_s\overline{X}_t\}$可以完全决定它的有限维分布,从而不仅可以在二阶矩的范围内,而且可以在分布的范围内讨论各种问题,得到不少更强的结果.

顺便指出,本节将假定所研究的正态过程都是实正态过程.如果所给的随机过程是一复正态过程,则类似地可以得到相应的结论.

22.4.1 正态过程的有限维分布

对正态过程$\{X_t,t\in T\}$任选 n 个时刻(理解 T 为时间集)t_i(n 也是任意的正整数),得到 n 个相应的随机变量,X_{t_i},$i=1,\cdots,n$.它们的联合分布密度为

定义 22.4.1 联合分布密度

$$f_X(x_{t_1},\cdots,x_{t_n})=\frac{1}{(2\pi)^{\frac{n}{2}}}\frac{1}{|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x_t-\mu_t)B^{-1}(x_t-\mu_t)^{\tau}\right\}$$

其中 $x_t=(x_{t_1},x_{t_2},\cdots,x_{t_n})$,$\mu_t=(\mu_{t_1},\mu_{t_2},\cdots,\mu_{t_n})\in R_n$,$\mu_{t_i}=E(x_{t_i})$,$i=1,2,\cdots,n$.

B 为 X_t 的协方差矩阵,为正定的,它的元素为

$$\begin{aligned}b_{kj}&=E\{(X_{t_k}-\mu_{t_k})(X_{t_j}-\mu_{t_j})\}\\&=R_X(t_k,t_j)-\mu_{t_k}\mu_{t_j}\end{aligned}$$

这里 $R_X(t_k,t_j)$为自相关函数.

如果$\{X_t,t\in T\}$为(宽)平稳的高斯过程,则 $\mu_{t_k}=\mu=$常数,$R_X(t_k,t_j)=R_X(t_k-t_j)$,即均值为常数,相关函数(或协方差函数)

仅与时间差 t_k-t_j 有关，而不再单独依赖于 t_k 和 t_j 本身.

记 $b_{kj}=R_X(t_k-t_j)-\mu^2=b(t_k-t_j)$，此时 n 元随机向量 $(X_{t_1},\cdots,X_{t_n})$ 的概率密度 f_X 所对应的特征函数为

公式 22.4.2 $\Phi(u_1,\cdots,u_n;t_1,\cdots,t_n)$

$$=\exp\left\{\mathrm{i}\left(\sum_{k=1}^{n}u_k\right)\mu-\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}b(t_k-t_j)u_ku_j\right\}.$$

如果把 n 个抽样时刻同时作一平移 h，即抽样时刻 $t_1+h,\cdots,t_n+h$，则平移后的 n 元随机向量的概率密度所对应的特征函数为

$$\Phi(u_1,\cdots,u_n;t_1+h,\cdots,t_n+h)$$

$$=\exp\left\{\mathrm{i}\left(\sum_{k=1}^{n}u_k\right)\mu-\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}b(t_k-t_j)u_ku_j\right\}$$

$$=\Phi(u_1,\cdots,u_n;t_1,\cdots,t_n).$$

因为分布函数由其特征函数唯一确定，故有

$$f(x_1,\cdots,x_n;t_1+h,\cdots,t_n+h)=f(x_1,\cdots,x_n;t_1,\cdots,t_n).$$

这就表明，当 $X_T=\{X_t,t\in T\}$ 是(宽)平衡高斯过程时，它的一切有限维分布都不随时间推移而改变，即 X_t 也是严格平稳随机过程.

22.4.2 正态过程的性质

由于正态变量间的独立与不相关等价，立得如 X_t 为正态过程，则过程独立⇔过程的协方差函数 $\Gamma(s,t)=0,s\neq t$.

1. 样本函数性质

设 X_t 为正态过程，均值函数，协方差函数分别为 $m(t)$ 和 $\Gamma(s,t)$.

性质 22.4.3 样本函数性质

(1) 如 X_T 可分，$m(t)\equiv 0$，$\exists\ c\geqslant 0,\varepsilon>0$，使 $\mathrm{Var}(X_s-X_t)\leqslant c|s-t|^{\varepsilon}$，则 X_T 样本函数以概率 1 连续.

(2) 如 X_T 可分，$m(t)$ 连续，$\exists\ c\geqslant 0,\varepsilon>0$，使 $|\Gamma(s,t)-\Gamma(s,s)|\leqslant c|t-s|^{\varepsilon}$，则 X_T 样本函数以概率 1 连续.

(3) 如 X_T 样本函数连续，$m(t)\equiv 0$，$\exists\ c>0, 0<\varepsilon<2$ 使方差 $V(X_s-X_t)\geqslant \sigma^2|s-t|^\varepsilon$，则样本函数以概率 1 几乎处处不可导.

2. 正态过程的不变性

设 X_T 为正态过程，且 T 为任意集(也称此种过程为正态系).

性质 22.4.4 正态过程的不变性(invariance of normal processes)

(1) 由 X_T 张成的线性集(即 X_T 中任何有限多个元素的线性组合之全体)，仍为正态系.

(2) X_T 在依概率收敛意义下的闭包(closure)仍为正交系.

(3) 设 $X_T^{(n)}=\{X_t^{(n)};t\in T\}$ 为正态过程，$n=1,2,\cdots$. 如 $X_t^{(n)}\xrightarrow{ms}X_t(n\to\infty)$，$\forall\ t\in T$. 则 $X_T=\{X_t;t\in T\}$ 如存在，则也是正态过程.

(4) 均方可微的正态过程 X_T，其均方导函数过程(称可导过程)$X_T'=\{X_t',t\in T\}$也是正态过程.

(5) 均方可积的正态过程 X_T，其不定积分过程$\int_a^t X_s\mathrm{d}s, t\in T$，$a(\in T)$ 固定，也是正态过程.

(6) 设 $h(t)$ 为一线性时不变系统(见定义 23.6.1)的冲激响应函数，X_T 为均方可积正态过程，则 $Y_t=\int_0^t X_s h(s)\mathrm{d}s, a\leqslant t\in T$，也是正态过程.

3. 正态过程微分与积分的特征函数

定理 22.4.5 设正态过程 X_T 均方可微，则$(X'(t_1),\cdots,X'(t_k))^{\mathrm{T}}$仍为正态，其特征函数为

$$\varphi_{X'}(u_1,\cdots,u_k;t_1,\cdots,t_k)$$
$$=\exp\left\{\mathrm{i}\sum_{j=1}^k u_j m'(t_j)-\frac{1}{2}\sum_{j,l=1}^k u_j u_l\frac{\partial^2\Gamma(t_j,t_l)}{\partial s\partial t}\right\}$$

其中

$$\frac{\partial^2\Gamma(t_j,t_l)}{\partial s\partial t}=\frac{\partial^2\Gamma(s,t)}{\partial s\partial t}\Big|_{s=t_j,t=t_l}.$$

定理 22.4.6 设正态过程 X_T 均方可积,$Y(t)=\int_a^t X_s\mathrm{d}s$ 则 $(Y(t_1),\cdots,Y(t_k))^{\mathrm{T}}$ 仍为正态过程,其特征函数为

$$\Phi_Y(u_1,\cdots,u_k;t_1,\cdots,t_k)$$

$$=\exp\left\{i\sum_{j=1}^k u_j\int_a^{t_j}m(s)\mathrm{d}s-\frac{1}{2}\sum_{j,l=1}^k u_ju_l\int_0^{t_j}\int_0^{t_l}\Gamma(s,t)\mathrm{d}s\mathrm{d}t\right\}.$$

22.4.3 正态过程作为输入通过非线性系统

设一非线性系统,它没有存储能量元件,即在 t 时的输出 $Y(t)$ 仅与该时刻 t 的输入 X_t 有关,其关系式为

$$Y(t)=g(X_t). \tag{22.2}$$

如果给定输入过程 X_t 的分布,且可利用(22.2)式作变量替换,得到输出过程 $Y(t)$ 的分布,即可写

$$P(Y(t)\leqslant y)=P(X_t\in A(y)),$$

其中 $A(y)=\{x:g(x)\in(-\infty,y]\}$.
如果上一等式对 y 的导数存在,对(22.2)式求导可得 $Y(t)$ 的概率密度.

如果输入过程的概率密度为 $f_{X_t}(x)$,且几乎处处连续,$Y(t)$ 和 X_t 的关系是一一对应关系,则输出过程的概率密度为

$$f_{Y_t}(y)=f_{X_t}(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|.$$

而且

$$E(Y(t))=\int_{-\infty}^{\infty}g(x)f_{X_t}(x)\mathrm{d}x,$$

$$E[(Y(t))^n]=\int_{-\infty}^{\infty}[g(x)]^nf_{X_t}(x)\mathrm{d}x,$$

$P(t)$ 在 t_1,t_2 时刻的相关函数为

$$
\begin{aligned}
R_Y(t_1,t_2) &= E(Y(t_1)Y(t_2)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)g(x_2)f_{X_{t_1}},X_{t_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2,
\end{aligned}
$$

式中 $f_{X_{t_1}},X_{t_2}(x_1,x_2)$代表在 t_1,t_2 时刻 X_{t_1},X_{t_2}的联合概率密度，它是二元正态的.

22.4.4 零交和阈交问题

设 X_T 是一随机过程，希望知道在某一时间间隔$[t_1,t_2]\subset T$中，$X(t)$的零点个数，或者更一般地，$X(t)$与某一水平线 $x=x_0$ 的交点个数，记为 $N(x_0;t_1,t_2)$. 这在信号处理、经济管理与系统控制中，是十分有用的. 由于 $X(t)$的随机性，因而 $N(x_0;t_1,t_2)$也是随机的，对固定 x_0,t_1,t_2，它是一随机变量，要求得到它的分布或其均值等数字特征.

这个问题一般称为**零交问题**($x_0=0$ 时)(zero crossing)和**阈交问题**(对一般的 x_0)(threshold crossing).

定理 22.4.7 设$\{X(t);t\in T\}$为均方可微的正态过程，$\delta(x)$是 δ 函数，则

$$
\begin{aligned}
N(x_0;t_1,t_2) &= \int_{t_1}^{t_2}|Y'(t)|\mathrm{d}t \\
&= \int_{t_1}^{t_2}|X_t'|\delta[X_t - X_0]\mathrm{d}t.
\end{aligned}
$$

$$
\begin{aligned}
E\{N(x_0;t_1,t_2)\} &= \int_{t_1}^{t_2}E\{|X_t'|\cdot\delta[X_t - X_0]\}\mathrm{d}t \\
&= \int_{t_1}^{t_2}\int_{-\infty}^{\infty}|x'|f(t,x_0;t,x')\mathrm{d}x'\mathrm{d}t,
\end{aligned}
$$

其中 $f(t,x_0;t,x')$为 X_t 与 X_t' 的联合分布密度.

记 $N_+(x_0;t_1,t_2)$和 $N-(x_0;t_1,t_2)$分别为与 x_0 向上和向下相交的个数，则有

$$
\begin{cases}
E(N_{+}) = \int_{t_1}^{t_2}\int_{-\infty} x' f(t, x_0; t, x') \mathrm{d}x\mathrm{d}t, \\
E(N_{-}) = -\int_{t_1}^{t_2}\int^{\infty} x' f(t, x_0; t, x') \mathrm{d}x\mathrm{d}t.
\end{cases}
$$

又记 $r(t, x_0) = \int_{-\infty}^{\infty} |x'| f(t, x_0; t, x') \mathrm{d}x'$，则 $r(t, x_0)$ 是 X_t 在 t 时与 x_0 相交的平均率. 特别在 X_T 是平稳时，$r(t, x_0)$ 与 t 无关，这时 $r(t, x_0) = r(x_0)$，它表示单位时间中，$X(t)$ 与 x_0 相交的次数. 利用下一定理，可求出 $r(x_0)$.

定理 22.4.8 设 X_T 是均方可微的实平稳过程，则 $EX_tX_t' = 0$. 又如 X_T 还是正态的，则 X_t 和 X_T' 独立.

定理 22.4.9 设 $\{X_t; t \in T\}$ 是实的均方可微的平稳正态过程，则

$$
r(X_0) = \frac{\sigma_{X'}}{\pi\sigma_X} \exp\left\{-\frac{1}{2}\frac{X_0^2}{\sigma_{X^2}}\right\},
$$

其中 $\sigma_X^2 = B(0)$，$\sigma_{X'}^2 = -B''(0)$，而 $B(\tau)$ 是 X_t 的均方差函数.

22.4.5 正态马尔可夫过程

定理 22.4.10 设 X_T 是正态过程，且 $E|X_t|^2 \neq 0$，则它同时又是马尔可夫过程的充要条件是其规范化的相关函数满足：

$$
R(t_1, t_3) = R(t_1, t_2) R(t_2, t_3), \quad t_1 < t_2 < t_3.
$$

其中

$$
R(s, t) = \begin{cases}
\dfrac{\Gamma(s, t)}{[\Gamma(s, s)\Gamma(t, t)]^{1/2}}, & \text{如 } \Gamma(s, s) \times \Gamma(t, t) > 0, \\
0, & \text{如 } \Gamma(t, t) = 0.
\end{cases}
$$

这里 $\Gamma(s, t)$ 是相关函数.

定理 22.4.11 设 $\{X_n, n = 0, \pm 1, \pm 2, \cdots\}$ 为正态分布、平稳实随机序列，且 $c(0) \neq 0$，则 $c(n) = a^n c(0), n \geqslant 0, |a| \leqslant 1$ 为 (X_n) 是马尔可夫过程的充分必要条件.

其中

$$c(s-t)=\mathrm{cov}\{X(s),X(t)\}.$$

定理 22.4.12 设 $X(t)$是一均方连续、平稳、实正态分布的随机过程,$c(\tau)$为其协方差函数,则 $c(\tau)=e^{a\tau}c(0),\tau\geqslant0,a<0$ 是该过程具有马尔可夫性的充分必要条件.

22.4.6 零初值布朗运动

定义 22.4.13 随机过程 $B_T=\{B_t,t\geqslant0\}$是一独立增量过程(定义 22.3.5),且 $B_t-B_s\sim N(0,t-s)$,$\forall\ t>s\geqslant0$ 时,则称 B_T 为**标准布朗运动**(standard Brownian motion),简记为 BM.

布朗运动又称维纳过程,详细讨论见第 26 章,这里只给出它与其它过程的关系.

(1) 零初值标准布朗运动是正态过程,均值函数为 0,而 $\Gamma(s,t)=\sigma^2\min(s,t)$,因此不是平稳过程.

(2) 零初值(标准)BM 是齐次独立增量过程,因而是齐次马尔可夫过程.

(3) BM 是均方连续的,由定理 22.2.21 及上述(1),它的样本函数处处不可微.

$\dfrac{\partial^2\Gamma}{\partial s\partial t}$不存在,故 BM_0 的均方导数不存在,但形式上可写 B_T 导过程 $B_T{}'$的协方差函数为

$$\Gamma_{B'}(s,t)=\frac{\partial^2}{\partial s\partial t}\Gamma(s,t)$$
$$=\delta(s-t),$$

其中 $\Gamma(s,t)$为 B_T 的协方差函数,$\delta(\cdot)$是 Dirac 函数(δ 函数). 当 $s\neq t$ 时,$\delta(s-t)$从而 $\Gamma_{B'}(s,t)$为 0,即 B_s'与 B_t'不相关. 这种过程 $B_T{}'$一般称为**白噪声**(white noise). 由于零初值 BM 是正态分布,故也称 $B_t',t\geqslant0$ 为**正态白噪声**(normal white noise).

当 $B_t - B_s \sim N(0, \sigma^2(t-s))$，$t > s \geqslant 0$ 时，写 $\Gamma_{B'}(s,t) = \sigma^2\delta(s-t)$，也称 B_T' 为正态白噪声.

22.5 泊松过程及其应用

22.5.1 计数过程与泊松过程定义

定义 22.5.1 一个随机过程 $\{N(t), t \geqslant 0\}$ 称为**计数过程**(counting process)，如果 $N(t)$ 表示在 $(0,t]$ 内事件 A 出现的总数. 因此，一个计数过程应当满足：

(1) $N(t) \geqslant 0$；

(2) $N(t)$ 是一非负整数；

(3) 如果 $s < t$，则 $N(s) \leqslant N(t)$；

(4) 对于 $s < t$，$N(t) - N(s)$ 表示时间间隔 $(s,t]$ 内事件 A 出现的次数.

一个计数过程称为独立增量的，如果不相交叠的时间间隔内事件出现的次数是相互独立的.

一个计数过程称为**平稳增量的**，如果事件出现的次数的分布仅依赖于时间间隔的长度，而与起始时间无关.

定义 22.5.2 一个计数过程 $\{N(t); t \geqslant 0\}$，称为**齐次泊松过程**(homogeneous Poisson process)，如果：

(1) $N(0) = 0$；

(2) $\{N(t), t \geqslant 0\}$ 是独立增量过程；

(3) $P\{N(h) = 1\} = \lambda h + o(h)$，$\lambda > 0$；

(4) $P\{N(h) \geqslant 2\} = o(h)$.

齐次泊松过程的另一定义及非齐次泊松过程见定义 27.2.1.

22.5.2 齐次泊松过程的几个问题

1. 泊松过程 $\{N(t); t \geqslant 0\}$ 在时间间隔 $(t_0, t_0 + t]$ 内事件出现 n

次的概率为

$$P\{[N(t_0+t)-N(t_0)]=n\}=\frac{(\lambda t)^n}{n!}e^{-\lambda t},\lambda>0,n=0,1,2,\cdots.$$

2. 各次到达时间间隔分布

记 X_1 为事件 A 第一次出现的时间间隔，一般地，对 $n\geqslant 1$，X_n 为事件第 n 次出现与第 $n-1$ 次出现的时间间隔，$\{X_n,n\geqslant 1\}$ 称为时间间隔序列，那么$\{X_n,n\geqslant 1\}$是一串独立同指数分布的随机变量，具有均值 $1/\lambda$.

3. 等待时间的分布

记 s_n 为从时间 $t=0$ 开始到达第 n 次事件出现所需的等待时间，则 s_n 的概率密度为 Γ 分布：

$$f_{s_n}(t)=\begin{cases}\lambda e^{-\lambda t}\dfrac{(\lambda t)^{n-1}}{(n-1)!} & t\geqslant 0;\\ 0, & t<0.\end{cases}$$

4. 到达时间的条件分布

给定 $N(t)=n$，$s_1,\cdots,s_n$ 的条件概率密度为

$$f(t_1,\cdots,t_n|N(t)=n)=n!/t^n,\quad 0<t_1<\cdots<t_n.$$

定理 22.5.3 给定 $N(t)=n$，则 n 次到达时间 $s_1,\cdots,s_n$ 与 n 个独立均匀分布于区间$(0,t)$内的随机变量的顺序统计量有相同的分布.

5. $\{N_1(t),t\geqslant 0\}$和$\{N_2(t),t\geqslant 0\}$为两个相互独立的泊松过程，它们在单位时间内事件出现的平均次数分别为 λ_1 和 λ_2. 记 $s_k^{(1)}$ 为第一过程中出现第 k 次事件的等待时间，$s_1^{(2)}$ 为第二过程中出现第一次事件的等待时间，那么，第一过程出现第 k 次事件先于第二过程出现第一次事件的概率为

$$P\{s_k^{(1)}<s_1^{(2)}\}=(\lambda_1/(\lambda_1+\lambda_2))^k.$$

22.5.3 滤过的泊松过程及应用

设一泊松分布的冲激脉冲串经过一线性时不变滤波器，则滤

波器输出是一随机过程$\{X_t, 0\leqslant t<\infty\}$：

$$X_t = \sum_{i=1}^{N(T)} h(t-U_i), \tag{22.3}$$

式中$h(t)$代表滤波器的冲激响应，U_i代表第i个冲激脉冲出现的时间，$N(T)$代表在$(0,T]$内进入到滤波器输入端冲激脉冲的个数，它服从泊松分布，即

$$P\{N(T)=k\} = \frac{(\lambda T)^k}{k!}e^{-\lambda T} \quad k=0,1,2,\cdots$$

这里λ代表单位时间内的平均脉冲数.

若在$(0,T]$内进入到滤波器输入端的脉冲数$N(T)$为k，则该k个脉冲出现的时间均为独立同分布的随机变量，且该随机变量均匀分布于$(0,T]$内，即

$$f_{(U_i|N(T)=k)}(u) = \begin{cases} \dfrac{1}{T}, & 0\leqslant U<T, \\ 0, & 否则. \end{cases}$$

称(22.3)式所表示的随机过程X_T为**滤过的**(filtered)**泊松过程**.

应用 22.5.4 温度限制的二极管中，由散弹效应引起的散弹噪声电流是滤过的泊松过程. 此时二极管板流为

$$I(t) = \sum_{j=1}^{N(T)} i(t-U_j), \quad t<T,$$

其中U_j是第j个电子的发射时刻，它是$[0,T)$上均匀分布的随机变量；函数$i(s)$为时刻s的板流，

$$i(s) = \begin{cases} 2q_e s/\tau_a^2, & 如\ 0\leqslant s\leqslant \tau_a, \\ 0, & 其它. \end{cases}$$

这里q_e为电子电荷，而τ_a为电子从阴极出发到达板极所需要的时间.

不难求得板流的统计特性：

$$EI(t) = \lambda q_e,$$

$$\Gamma(t,t+\tau)=\lambda\int_0^t i(t)i(t+\tau)\mathrm{d}t,\quad \tau>0,$$

$$V(I(t))=\lambda\int_0^\tau i^2(t)\mathrm{d}t,$$

其中 λ 为单位时间内发射的平均电子数.

22.5.4 韦布尔过程及其应用

1. 韦布尔过程

定义 22.5.5 称一强度函数为 $\lambda(t)=\lambda\beta t^{\beta-1}$ 的(非齐次)泊松过程(定义 27.2.1)为**韦布尔过程**(Weibull process),这里 λ, $\beta>0$, $t\geqslant 0$. 称

$$\Lambda(t)=\int_0^t\lambda(u)\mathrm{d}u=\lambda t^\beta,\quad t\geqslant 0$$

为累积强度函数.

上述韦布尔过程 $\{N(t),t\geqslant 0\}$ 的事件首次发生(或第一个质点到达)的时间 X_1,有韦布尔分布. 事实上

$$p(X_1>t)=p(N(t)=0)=\mathrm{e}^{-\Lambda(t)}=\exp(-\lambda t^\beta),\quad t\geqslant 0.$$

但第二、第三次……发生的相邻间隔时间都不再是韦布尔分布. 相邻间隔时间,当 $0<\beta<1$ 时呈现变大趋势,而 $\beta>1$ 时则呈变小趋势. $\beta=1$ 化为齐次泊松过程.

2. 韦布尔过程模型的统计分析

1) 定数观察情形

对事先给定的 n,观察到第 n 个事件发生的时刻 t_n 终止,可获得数据

$$0\leqslant t_1\leqslant t_2\leqslant\cdots\leqslant t_n.$$

用韦布尔过程作模型(例如可靠性模型,计数中的"事件"解释为系统故障)时,要估计未知参数 λ 和 β.

该 $T_1,T_2,\cdots,T_n$ 为头 n 个事件发生的时刻,则其联合密度为

$$f(t_1,t_2,\cdots,t_n)=\exp(-\lambda t_n^{\beta})\prod_{i=1}^{n}(\lambda\beta t_i^{\beta-1}).$$

由$\dfrac{\partial f}{\partial\lambda}=0$，$\dfrac{\partial f}{\partial\beta}=0$解得$\lambda,\beta$的最大似然估计(MLE)为

$$\begin{cases}\hat{\beta}=n\left[\sum_{i=1}^{n}\ln(T_n/T_i)\right]^{-1},\\ \hat{\lambda}=n/(T_n\hat{\beta}).\end{cases}$$

定理 22.5.6 设$T_i,i=\overline{1,n}$为韦布尔过程头n个事件发生的时刻，则当$T_n=t_n$已知时

(1) $2n\beta/\hat{\beta}\sim\chi^2_{2n-2}$

(2) $2\lambda T_n^{\beta}\sim\chi^2_{2n}$

(3) $\hat{\beta}\ln(\lambda^{\frac{1}{\beta}}T_n)$与$2n\ln(Y/2)/Z$同分布，其中$Y,Z$相互独立，分别有$\chi^2_{2n}$及$\chi^2_{2n-2}$分布.

由此定理，β的置信度为$1-\alpha$的置信区间为

$$\left[\frac{\hat{\beta}}{2n}\chi^2_{2n-2}\left(\frac{\alpha}{2}\right),\ \frac{\hat{\beta}}{2n}\chi^2_{2n-2}\left(1-\frac{\alpha}{2}\right)\right].$$

若$2n\ln(Y/2)/Z$的α分位数记为$h(\alpha)$(可由计算机给出)，则$\lambda^{1/\beta}$的$1-\alpha$双侧置信区间为

$$\left[\frac{1}{T_n}\exp\left(\frac{1}{\hat{\beta}}h\left(\frac{\alpha}{2}\right)\right),\ \frac{1}{T_n}\exp\left(\frac{1}{\hat{\beta}}h\left(1-\frac{\alpha}{2}\right)\right)\right].$$

2) 定时观察情形

观察在指定时刻t终止，设有数据$0\leqslant t_1\leqslant t_2\leqslant\cdots\leqslant t_n\leqslant t$. 仿 1) 可得$\beta$与$\lambda$的 MLE 为

$$\begin{cases}\hat{\beta}=n\left[\sum_{i=1}^{n}\ln(t/T_i)\right]^{-1},\\ \hat{\lambda}=n/(t\hat{\beta}).\end{cases}$$

定理 22.5.7 在$N(t)=n$条件下，有$2n\beta/\hat{\beta}\sim\chi^2_{2n}$.

由此可得β的$1-\alpha$置信区间为

$$\left[\frac{\hat{\beta}}{2n}\chi^2_{2n}\left(\frac{\alpha}{2}\right),\ \frac{\hat{\beta}}{2n}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right].$$

3. 在可靠性增长模型中的应用

假定一系统开发阶段如下进行：系统研制成功后进行试验，直到出现一个故障为止. 寻找原因并改进系统，消除引起此种故障发生的故障模式，然后再做试验、再修改，直到试验达到预定可靠性要求或已满足规定的试验期长度为止. 这样，系统的可靠性在此开发阶段是不断增长的. 若用 $N(t)$ 表 $(0,t]$ 中系统故障次数，而维修改进的时间作 0 处理. 许多大型复杂系统的开发阶段数据表明，$(t,EN(t))$ 间有关系 $\ln EN(t)=\beta\ln t+\alpha,\beta>0$. 令 $\lambda=e^{\alpha}$，则 $EN(t)=\lambda t^{\beta}$，即 $\Lambda(t)=\lambda t^{\beta}$，$\{N_t,t\geqslant0\}$ 为 Weibull 过程.

当系统在时刻 t_0 定型投产，则可认为 t_0 后系统不再作设计上的改进，而有常数失效率 $\lambda(t_0)=\lambda\beta t_0^{\beta-1}$. 寿命有指数分布，平均寿命为 $t_0^{1-\beta}/(\lambda\beta)$. 开发阶段，$\beta<1$，视为增长参数.

注 复合泊松过程(compound Poisson process)请读节 27.6 及[30]，独立增量过程在排队论中应用见第 32 章及[70].

23 平稳随机过程与时间序列

23.1 平稳过程的定义

平稳过程(stationary process)是对于时间的推移具有某种平稳性的随机过程的总称.它分为**强平稳过程**(stroctly stationary process)和**弱平稳过程**(weakly stationary process)(或宽平稳过程).

平稳过程应用广泛.许多领域如无线电技术、自动控制等方面,随着时间的演变,在工作条件基本稳定的情况下,可以认为是平稳过程.

23.1.1 弱平稳过程

定义 23.1.1 设 $X_T=\{X(t),t\in T\}$ 是定义在概率空间 $(\Omega,\mathscr{F},P)$ 上的复值随机过程,T 为实数集或其子集,若 X_T 满足

(1) 对任意 $t\in T$, $E|X(t)|^2<\infty$,

(2) 对任意 $t\in T$, $t+\tau\in T$,

$$E(X(t+\tau)\overline{X(t)})=B(\tau),$$

则称 X_T 为**弱平稳过程**,或**宽(广义)平稳随机过程**,简称**平稳过程**. $B(\tau)$称为 X_T 的**相关函数**(correlation function).

定义 23.1.2 设 $X_T=\{X(t),t\in T\}$是平稳过程,若 T 是离散集合,称 X_T 为**平稳序列**(stationary series),若 T 还是时间集合,称 X_T 为**平稳时间序列**(stationary time series).

23.1.2 相关函数

性质 23.1.3 相关函数 $B(\tau)$的性质

(1) $B(0)\geqslant 0$

(2) $|B(\tau)|\leqslant B(0)$

(3) $B(-\tau)=\overline{B(\tau)}$

(4) 非负定性(non-negative definete)对任意自然数 n,任意复数 $a_1,a_2,\cdots,a_n$,任意 $t_1,t_2,\cdots,t_n\in T$,有

$$\sum_{i,j=1}^{n} B(t_i - t_j)a_i\bar{a}_j \geqslant 0.$$

定理 23.1.4 为使函数 $B(\tau)$为某弱平稳过程的相关函数,充分必要条件是它具有(23.1.3)中的性质(4).

定理 23.1.5 对于平稳过程$\{X(t),t\in T\}$,下列四个条件等价:

(1) 过程为均方连续;

(2) 过程在 $t=0$ 点均方连续;

(3) $B(\tau)$在 R_1 上连续;

(4) $B(\tau)$在 $\tau=0$ 点连续.

本文以下恒设 X_T 为均方连续的,且不失一般性,可设 $EX_t=m=0$,否则可考虑 $X_t'=X_t-EX_t$, $t\in T$.

例 23.1.6 若 $X=\{X(n),n\in N\}$为实的互不相关随机变量序列,且 $E(X(n))=0,D(X(n))=\sigma^2,n\in N,N$ 为整数列. 则$\forall$ m, $n\in N$,

$$E(X(n)X(m)) = \begin{cases}\sigma^2, & m = n;\\ 0, & m \neq n.\end{cases} \tag{23.1}$$

即相关函数只与 $n-m$ 有关,所以它是平稳过程.

在工程技术上,常把上述随机变量序列称为“**白噪声**”(white noise). 它普遍地存在于各种波动现象中(如电流的波动,电子发射的波动和通讯设备各部分电流或电压的波动等).

例 23.1.7 设随机变量序列$\{X(n),n\in N\}$满足规范化条件 $E|X(n)|^2=1$ 及条件 $E(X(n+\tau)\overline{X(n)})=0,\tau\neq 0,\tau\in N$,并用均

方收敛定义

$$x_n = \sum_{i=-\infty}^{\infty} a_i X(n-i), \quad 其中 \sum_{i=-\infty}^{\infty} |a_i|^2 < \infty;$$

$$y_n = \sum_{i=0}^{\infty} b_i X(n-i), \quad 其中 \sum_{i=0}^{\infty} |b_i|^2 < \infty;$$

$$z_n = \sum_{i=0}^{m} c_i X(n-i), \quad m \text{ 固定};$$

a_i, b_i, c_i 均为复数，则 $\{X(n), n\in N\}$，$\{y(n), n\in N\}$，$\{z_n, n\in N\}$ 均为弱平稳序列，它们的相关函数分别为

$$B_x(\tau) = \sum_{i=-\infty}^{\infty} a_i \bar{a}_{i-\tau},$$

$$B_y(\tau) = \sum_{\substack{i\neq 0 \\ i-\tau\geqslant 0}} b_i \bar{b}_{i-\tau},$$

$$B_z(\tau) = \sum_{\substack{m\geqslant i>0 \\ m\geqslant i-\tau\geqslant 0}} c_i \bar{c}_{i-\tau},$$

并称 x_n, y_n, z_n 为序列的**滑动和**.

例 23.1.8 随机电报信号　考虑 $\zeta_t = X(-1)^{\mu_t}, t\geqslant 0$，对每个 t，ζ_t 的取值仅有 0，1 两种状态，每种状态持续时间是随机的(参见图 23.1).

(1) r.v. $X \sim \begin{pmatrix} -1 & 1, \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

(2) 在 t 时间内，波形变化次数 μ_t 服从参数为 λt 泊松分布，对固定的 λ，它是泊松流. 故

$$P(\mu_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0,1,2,\cdots$$

其中 $\lambda>0, t>0$.

(3) 对任意 $t, t>0$，X 与 μ_t 独立，则 $\{\xi_t\}$ 是平稳过程.

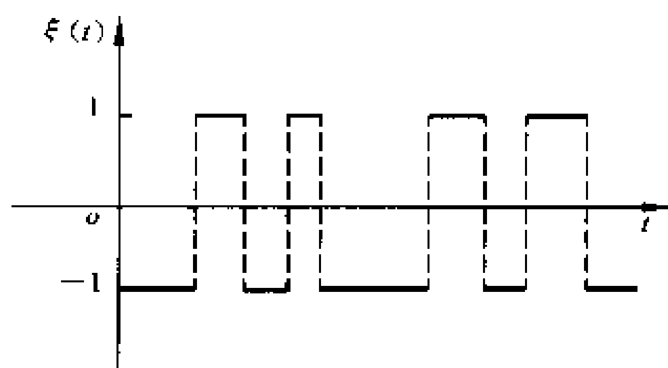


图 23.1

23.2 相关函数和平稳过程的谱分解

23.2.1 相关函数的谱分解

1. 谱分解定理

定理 23.2.1 波赫拿-辛钦(Bochner-Хинчин)定理 设 $X_T=\{X(t),t\in T\}$是均方连续的均值为 0 的平稳过程. $B(t)$是它的相关函数.

(1) 如果 X_T 是离散时间数值过程,则

$$B(t)=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}\lambda t}\mathrm{d}F(\lambda), \tag{23.2}$$

其中 $F(\lambda)$是有界右连续、非负不减函数且不计常数之差外,唯一决定. 如要求 $F(-\pi)=0$,则 $F(\lambda)$唯一.

(2) 如果 X_T 是离散时间向量过程,则 $B(t)$有公式(23.2)的表现,其中 $F(\lambda)$是矩阵,对于 $\lambda_1\geqslant\lambda_2$, $F(\lambda_1)-F(\lambda_2)$是埃尔米特(Hermite)的和非负定的,如果假设 $F(\lambda)$(按元收敛意义下)右连续,且 $F(-\pi)=O$,则 $F(\lambda)$唯一.

(3) 如果 X_T 是连续时间数值过程,则

$$B(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\lambda t}\mathrm{d}F(\lambda), \tag{23.3}$$

其中 $F(\lambda)$的意义与情形(1)相同，只是 $F(-\infty)=O$.

(4) 如果 X_T 是连续时间向量过程，则 $B(t)$有(23.3)公式的表现. 其中矩阵 $F(\lambda)$的含义与情形(2)相同，但矩阵 $F(-\infty)=O$.

此定理称为相关函数的**谱分解定理**(spectral decomposition theorem). $F(\lambda)$称为过程 $X_T=\{X(t),t\in T\}$的**谱函数**(spectral function)，而谱函数 $F(\lambda)$产生的测度称为**谱测度**(spectral measure).

下列等式成立

$$B(0)=\begin{cases}F(\pi), & \text{离散时间;}\\ F(\infty), & \text{连续时间.}\end{cases}$$

2. 谱密度(spectral density)

定理 23.2.2 均方可积平稳过程的谱函数 $F(\lambda)$有勒贝格分解

$$F(\lambda)=F_1(\lambda)+F_2(\lambda)+F_3(\lambda).$$

这里 $F_1(\lambda)$关于勒贝格测度绝对连续，即存在可积函数 $f(\lambda)$，使

$$F_1(\lambda)=\begin{cases}\displaystyle\int_{-\pi}^{\lambda}f(\lambda)\mathrm{d}\lambda, & \text{离散时间;}\\ \displaystyle\int_{-\infty}^{\lambda}f(\lambda)\mathrm{d}\lambda, & \text{连续时间.}\end{cases}$$

$F_2(\lambda)$只可能在有穷或可列个 λ 点集合上跳跃地变化，$F_3(\lambda)$连续，并且关于勒贝格测度几乎处处有等于 0 的导数.

定义 23.2.3 谱函数(矩阵)$F(\lambda)$的绝对连续分量的导数 $F_1'(\lambda)=f(\lambda)$叫**谱密度函数**(矩阵)(spectral density function).

设 $X_T=\{X(t),t\in T\}$是向量过程，其谱密度函数 $f(\lambda)$绝对连续，如果它的谱密度矩阵 $f(\lambda)=F'(\lambda)$的秩等于 r，$r=\overline{1,k}$，其中 k 是过程的维数，则称 X_T 是秩为 r 的过程. 若 $\det f(\lambda)\neq 0$(几乎对一切 λ)，则称 X_T 为**最大秩过程**(maximan rank process).

当谱密度 $f(\lambda)$存在，$\int|B(t)|\mathrm{d}t<\infty$，相关函数 $B(t)$的谱分

解式为

$$B(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\lambda t}\mathrm{d}F(\lambda)=\begin{cases}\displaystyle\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\lambda t}f(\lambda)\mathrm{d}\lambda, & \text{连续时间;}\\ \displaystyle\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}n\lambda}f(\lambda)\mathrm{d}\lambda, & \text{离散时间;}\end{cases}\tag{23.4}$$

作 Fourier 逆变换，可知对连续情况

$$f(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\lambda t}B(t)\mathrm{d}t.\tag{23.5}$$

在实情况下

$$f(\lambda)=\frac{2}{\pi}\int_{0}^{\infty}\cos\lambda tB(t)\mathrm{d}t,\qquad \lambda\geqslant 0.\tag{23.6}$$

对离散情况，当 $\sum|B(n)|^2<\infty$ 时，有

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)\mathrm{e}^{-\mathrm{i}n\lambda}.\tag{23.7}$$

在实情况下

$$f(\lambda)=\frac{1}{\pi}\sum_{n=0}^{\infty}B(n)\cos n\lambda.\tag{23.8}$$

例 23.2.4 白噪声序列(例 23.1.6)的相关函数求得

$$B(n)=\begin{cases}\sigma^2, & n=0;\\ 0, & n\neq 0.\end{cases}$$

则由式(23.7)谱密度函数

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)\mathrm{e}^{-\mathrm{i}n\lambda}=\frac{\sigma^2}{2\pi},\qquad -\pi\leqslant\lambda\leqslant\pi.$$

而谱函数为 $F(\lambda)=\int_{-\pi}^{\lambda}f(\omega)\mathrm{d}\omega=\frac{\sigma^2}{2\pi}(\lambda+\pi)$.

例 23.2.5 若 $X(n)$ 为白噪声序列，令

$$Y(n)=\sum_{k=-\infty}^{\infty}C_{n-k}X(k),$$

则 $Y(n)$ 的相关函数与谱密度函数分别为

$$B(n)=\sigma^2\Big(\sum_{k=-\infty}^{\infty}C_{n+k}\bar{C}_k\Big),$$

$$f(\lambda)=\frac{1}{2\pi}\left|\sum_{k=-\infty}^{\infty}C_k e^{-i\lambda k}\right|^2.$$

例 23.2.6 设$\{\xi(t),t\in T\}$是随机电报信号，其相关函数为

$$B(\tau)=\frac{1}{4}e^{-2\lambda|\tau|},\qquad -\infty<\tau<+\infty.$$

如图 23.2 所示.

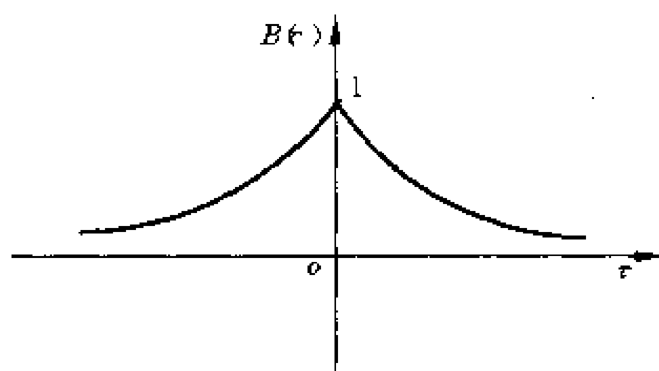


图 23.2

谱密度函数 $f(\omega)=\frac{1}{2\pi}\ \frac{\lambda_0^2}{4\lambda_0^2+\omega^2}$，$\quad -\infty<\omega<+\infty$，

式中 λ_0 是泊松分布的均值.

公式 23.2.7 几种重要随机过程的相关函数及相应的谱密度函数如表 23.1.

23.2.2 平稳过程的谱表现

定理 23.2.8 任何以 $F(\lambda)$ 为谱函数的均方连续平稳过程 $X_T=\{X(t),t\in T\}$，都有**谱表现**（或**谱分解**）(spectral representation)

$$X(t)=\int_{\Lambda}e^{i\lambda t}d\zeta(\lambda),\tag{23.9}$$

其中

$$\Lambda=\begin{cases}[-\pi,\pi], & \text{离散时间;}\\(-\infty,+\infty), & \text{连续时间.}\end{cases}$$

表 23.1

相关函数	谱密度
$\sigma^2 e^{-\alpha\|\tau\|} \quad (\alpha>0)$	$\dfrac{\sigma^2\alpha}{\pi(\alpha^2+\lambda^2)}$
$\sigma^2 e^{-\alpha\tau^2} \quad (\alpha>0)$	$\dfrac{\sigma^2}{2\sqrt{\pi\alpha}}e^{-\lambda^2/4\alpha}$
$\sigma^2 e^{-\alpha\tau^2}\cos\beta\tau \quad (\alpha>0)$	$\dfrac{\sigma^2}{4\sqrt{\pi\alpha}}\left\{e^{-\frac{(\lambda+\beta)^2}{4\alpha}}+e^{-\frac{(\lambda-\beta)^2}{4\alpha}}\right\}$
$\sigma^2 e^{-\alpha\|\tau\|}\cos\beta\tau \quad (\alpha>0)$	$\dfrac{\sigma^2\alpha}{2\pi}\left[\dfrac{1}{(\lambda-\beta)^2+\alpha^2}+\dfrac{1}{(\lambda+\beta)^2+\alpha^2}\right]$
$2a\dfrac{\sin b\tau}{\tau}^{*} \quad (a,b>0)$	$f(\lambda)=\begin{cases}a & \|\lambda\|\leqslant b\\ 0 & \|\lambda\|>b\end{cases}$
$Ae^{-\alpha\|\tau\|}(1+\alpha\|\tau\|)$	$\dfrac{2A\alpha^3}{\pi(\lambda^2+\alpha^2)^2}$
$B(\tau)=\begin{cases}\sigma^2(1-\|\tau\|), & \|\tau\|\leqslant 1\\ 0, & \|\tau\|>1\end{cases}$	$\dfrac{\sigma^2}{2\pi}\left(\sin\dfrac{\lambda}{2}\Big/\dfrac{\lambda}{2}\right)^2$
$Ae^{-\alpha\|\tau\|}\left(1+\alpha\|\tau\|+\dfrac{1}{3}\alpha^2\tau^2\right)$	$\dfrac{8A\alpha^5}{3\pi(\lambda^2+\alpha^2)^3}$

注：* 相关函数 $B(\tau)=2a\dfrac{\sin b\tau}{\tau}$ 是工程中常常用来近似拟合许多相关函数的一种函数形式，需注意该式是可积函数，但却不是绝对可积的.

积分视为黎曼-斯蒂尔切斯(Riemann-Stieltjes 简记 R-S)积分定义中和的序列的均方极限；$\zeta(\lambda)$为正交增量过程，满足条件

$$E(\zeta(\lambda))=0,\quad E(\mathrm{d}\zeta(\lambda)\mathrm{d}\,\overline{\zeta(\lambda)})=\mathrm{d}F(\lambda).$$

如果假设过程 $\zeta(\lambda)$在均方意义下右连续，则 $\zeta(\lambda)$唯一确定(精确

到 Ω 的 0 概率子集).

定义 23.2.9 公式(23.9)中的过程 $\zeta(\lambda)$ 称为对应于平稳过程 X_T 的**谱过程**(spectral process).

按照正交增量过程的均方积分定义(23.9)的积分是下列和式的极限

$$\lim_{N\to\infty}\ \mathrm{l.i.m}_{\Delta N\to 0}\sum_{j=0}^{N-1}\mathrm{e}^{it\lambda_j'}[\zeta(\lambda_{j+1})-\zeta(\lambda_j)],$$

其中 $\mathrm{e}^{it\lambda_j'}$ 是以 λ_j' 为频率的简谐振动,而 $[\zeta(\lambda_{j+1})-\zeta(\lambda_j)]$ 就是这个简谐振动的振幅,平稳过程的谱表现说明它是由不可数多个、不同频率的具有随机振幅的简谐波的叠加. 由于 $\zeta(\lambda)$ 是正交增量过程,所以这些简谐波的振幅是互不相关的. 平稳过程就是这些物理现象的数学描述.

平稳过程的谱表现首先由 А. Н. Колмогоров. 于 1941 年提出任何平稳过程这种展开式的可能性. 这一结果以希尔伯特空间 H 的术语陈述,且是从某些运算子谱理论中引导出来的.

23.3 平稳过程及其轨道的解析性质

23.3.1 平稳过程的均方连续性和可微性

设 $\{X(t), t\in T\}$ 是连续时间数值过程, $B(t)$ 和 $F(\lambda)$ 分别为它的相关函数和谱函数,平稳过程作为二阶矩过程的特例,有下述判别准则.

定理 23.3.1 平稳过程 $X(t)$ 均方连续的充分必要条件是它的相关函数 $B(t)$ 在点 $t=0$ 连续.

定理 23.3.2 平稳过程 $X(t)$ 有 m 阶均方导数的充分必要条件是它的相关函数 $B(t)$ 在点 $t=0$ 有 m 阶导数,或存在有穷 $2m$ 阶谱 $S_{2m}=\int_{-\infty}^{\infty}\lambda^{2m}\mathrm{d}F(\lambda)$.

若 $S_2=\int_{-\infty}^{\infty}\lambda^2 dF(\lambda)<\infty$，则过程 $y(t)=\{(X'(t),X(t)),t\in T\}$ 是宽平稳过程，它的矩阵相关函数 $B_y(t)$ 为

$$B_y(t)=\begin{bmatrix}\int_{-\infty}^{\infty}e^{i\lambda t}\lambda^2 dF(\lambda) & \int_{-\infty}^{\infty}e^{i\lambda t}e^{\lambda}dF(\lambda)\\ \int_{-\infty}^{\infty}e^{i\lambda t}e^{\lambda}dF(\lambda) & \int_{-\infty}^{\infty}e^{i\lambda t}\lambda^2 dF(\lambda)\end{bmatrix}$$

23.3.2 轨道的解析性质

定理 23.3.3 如果对某个 $q>3$，当 $t\to 0$ 时，有 $2B(0)-B(t)-B(-t)=0\left(\frac{|t|}{(\ln|t|)^q}\right)$，则以 $B(t)$ 为相关函数的过程等价于某一过程，其轨道在任何有穷区间上以概率 1 连续. 特别，如果 $B(t)$ 在点 $t=0$ 有二阶导数，则条件（定理 23.3.3）成立.

定理 23.3.4 如果对某个 $q>3$，当 $t\to 0$ 时有

$$6B(0)-4B(t)-4B(-t)+B(2t)+B(-2t)$$
$$=o\left(\frac{t^3}{|\ln|t||^q}\right),$$

则以 $B(t)$ 为相关函数的过程 $\{X(t),t\in T\}$ 等价于某一过程，其轨道以概率 1 可微. 特别，如果 $B(t)$ 在 $t=0$ 有 4 阶导数，则定理(23.3.4)成立.

定理 23.3.5 如果平稳过程的谱函数 $F(\lambda)$ 只在有穷区间上变化，则存在过程的等价过程，而且后者的轨道以概率 1 解析.

23.3.3 遍历性

研究随机过程的统计特性，一般要知道过程的 n 维分布函数或 n 维概率密度函数，这在实际工作中往往办不到. 能否从一个时间范围内观察到的一个样本函数作为提取这个过程的数字特征的充分依据？显然，一般情况下办不到. 如果一个过程从它的任一个样本函数，作某种时间上的平均就能得到此过程的相应的统计特

性，则称此过程有遍历性(各态历经性)．例如，若过程 $X=\{X(t), t\in T\}$ 的任一个样本函数 $x(t), t\in T$，平均值 $\sum\limits_{t\in T} x(t)/T = EX(t)$，则称 X 有均值遍历性(Ergodic)．这条件也可写成 $\sum\limits_{t\in T} X(t)/T = EX(t)$，a.s．有各态历经性(遍历性)的随机过程．

定义 23.3.6 (1)设 $\{X(t), t\in T\}$ 是均方连续、零均值平稳过程，$B(t)$ 和 $F(\lambda)$ 分别是它的相关函数和谱函数，$X(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(\lambda)$ 是它的谱表现，称随机变量

$$\hat{X}_s = \begin{cases} \dfrac{1}{s}\sum\limits_{t=0}^{s-1} X(t), & T=\{0,1,2,\cdots\},\ s\text{ 是正整数；} \\ \dfrac{1}{2s+1}\sum\limits_{t=-s}^{s} X(t), & T=\{0,\pm 1,\pm 2,\cdots\},\ s\text{ 是正整数；} \\ \dfrac{1}{s}\int_0^s X(t)dt, & T=(0,+\infty),\ s>0; \\ \dfrac{1}{2s}\int_{-s}^{s} X(t)dt, & T=(-\infty,+\infty),\ s>0. \end{cases}$$

为 X 的**时间均值**.

(2) 设

$$\hat{B}_s = \begin{cases} \dfrac{1}{s}\sum\limits_{t=0}^{s-1} B(t), & T=\{0,1,2,\cdots\},\ s\text{ 是正整数；} \\ \dfrac{1}{2s+1}\sum\limits_{t=-s}^{s} B(t), & T=\{0,\pm 1,\pm 2,\cdots\},\ s\text{ 是正整数；} \\ \dfrac{1}{s}\int_0^s B(t)dt, & T=(0,+\infty),\ s>0; \\ \dfrac{1}{2s}\int_{-s}^{s} B(t)dt, & T=(-\infty,+\infty),\ s>0. \end{cases}$$

称 $\hat{B}_s$ 为 $X(t)$ 的时间**自相关函数**(autocorrelation function).

定义 23.3.7 设 $\{X(t), t\in T\}$ 为均方连续、均值为零的平稳

过程，当 $s\to\infty$ 时，时间均值 $\hat{X}_s$ 的均方极限存在，称此过程具有**遍历性**.

定理 23.3.8 设 X_T 为均方连续，均值为零的平稳过程，则

$$\underset{s\to\infty}{\mathrm{L.i.m}}\hat{X}_s = \zeta(0) - \zeta(0^-), \underset{s\to\infty}{\mathrm{l.i.m}}\,\hat{B}_s = F(0) - F(0^-)$$

定理 23.3.9 $\underset{s\to\infty}{\mathrm{L.i.m}}\hat{X}_s = E(X(t)) = 0$ 的充要条件是谱函数 $F(\lambda)$ 在 0 点连续.

这样，如 $F(\lambda)$ 在 0 点连续，则 $X(t)$ 的时间平均，均方收敛于 $X(t)$ 的统计平均 $E(X(t))$，工程中常常利用这种遍历转换技术测量具有各态历经的随机过程的均值.

香农(Shannon)提出的**采样定理**(sampling theorem) 对于通信、控制理论起着十分重要的作用. 在运用计算机对随时间连续变化的信号 $x(t)$ 进行处理时，首先必须对信号进行离散的"采样". 一般每隔定长时间 Δ 后对 $x(t)$ 进行观察，获得 $x(t)$ 在 $t=k\Delta$ 时的数值 $x_k = x(k\Delta)$. Δ 愈小，同样时间内采样点愈多，就能更好反映 $x(t)$ 的特性. 但 Δ 太小，采样设备负担重，计算量也过大. 对于平稳过程，有下述定理.

定理 23.3.10 若 $X_T = \{X(t), t\in T\}$ 是均方连续的平稳过程，又 X_T 的谱函数 $F(\lambda)$，对某个 $B>0$ 满足

$$\int_{|\lambda|\geqslant 2\pi B} \mathrm{d}F(\lambda) = 0, \tag{23.10}$$

则

$$X(t) = \sum_{k=-\infty}^{\infty} X(k\Delta)\frac{\sin\frac{\pi}{\Delta}(t-k\Delta)}{\frac{\pi}{\Delta}(t-k\Delta)}, \quad \Delta = \frac{1}{2B}, \tag{23.11}$$

上述级数对每个 t 按均方意义下是收敛的.

实用中常遇到的情况是 X_T 有谱密度 $f(\lambda)$，这时相应于定理 23.3.9 的情况是

$$f(\lambda) = 0, \quad |\lambda| \geqslant \frac{1}{2B}. \tag{23.12}$$

(23.10)与(23.11)用来选择 $B=\frac{\lambda_0}{2\pi}$. 过程 X_T 不包含高于 λ_0 的频率成份. 因此称 λ_0 为**截止频率**(truncated frequency)或奈奎斯特(Nyquist)频率.

利用采样定理,对时间连续信号,平稳过程 $X_T=\{X(t),t\in T\}$,$T=(-\infty,+\infty)$,采样值 $X=\{X(n\Delta),n=0,\pm 1,\pm 2,\cdots\}$是一个平稳时间序列. 实用中采样间隔常取 $\Delta/2$ 或 $\Delta/3$.

23.4 平稳正态过程

23.4.1 定义

定义 23.4.1 设 $X_T=\{X(t),t\in T\}$是平稳过程,对任意的 $n\geqslant 1$ 和任意一组$(t_k\in T,k=\overline{1,n})$,向量$\{X(t_1),X(t_2),\cdots,X(t_n)\}$有多维正态分布,则称 X_T 为**平稳正态过程**(stationary normal process),或**平稳高斯过程**.

平稳正态过程完全取决于自己的数学期望和协方差函数. 反之,每一个函数 $m(t)=$常数和满足$|B(t)|\leqslant B(0)$及 $B(t),t\in T$ 连续,则可决定某一平稳正态过程.

正态过程是二阶矩过程的一个重要子类,它在实际中有着广泛的应用.

23.4.2 平稳正态过程的若干结果

定理 23.4.2 对可分的平稳正态过程$\{X(t),t\in T\}$,如果存在常数 $\varepsilon>0,c\geqslant 0$,使

$$|B(\tau)-B(0)| \leqslant c|\tau|^{\varepsilon},$$

则它的样本函数以概率 1 连续.

定理 23.4.3 设$\{X(t), t\in R_1\}$是均方连续的博雷尔可测的平稳正态过程，则它为遍历的充分必要条件是谱函数$F(\lambda)$是连续函数.

定理 23.4.4 设$\{X(t)\}$为平稳正态过程，则$\{X(t)\}$是马尔可夫过程的充要条件为对任意$\tau_1>0, \tau_2>0$有

$$R(\tau_1+\tau_2)=R(\tau_1)R(\tau_2),$$

其中$R(\tau)=E\{X(t+\tau)X(t)\}/E\{[X(t)]^2\}=\dfrac{B(\tau)}{B(0)}$.

定理 23.4.5 设$\{X(t)\}$为平稳正态过程，则$\{X(t)\}$是均方连续的马尔可夫过程的充要条件为其相关函数

$$B(\tau)=\mathrm{e}^{-a|\tau|}B(0),\quad a\geqslant 0 \text{ 常数}. \tag{23.13}$$

对$\{X(n), n=0,\pm 1,\cdots\}$，则(23.13)式改为

$$B(n)=a^{|n|}B(0),\quad |a|\leqslant 1. \tag{23.14}$$

23.5 强平稳过程

本节以下，为简洁，也常用小写字母表示随机过程.

定义 23.5.1 设$T=[0,+\infty)$，考虑概率空间$(\Omega,\mathscr{F},P)$及其定义于其上的实值过程$X(\omega)=\{x(t,\omega), t\in T\}$，对任意常数$s\geqslant 0$，任意正整数$n$，任意$t_i\in T, c_i\in R_1, i=\overline{1,n}$，满足

$$P(x(t_i)\leqslant c_i, i=\overline{1,n})=P(x(s+t_i)\leqslant c_i, i=\overline{1,n})$$

称$X(\omega)$为**强平稳过程**(strongly stationary process).

若对每一固定$s>0$，定义随机过程$X_s(\omega)=\{x(s+t,\omega), t\in T\}$.

定理 23.5.2 为使$X(\omega)$是强平稳过程，当且仅当下列诸条件之一满足：

(1) 对任意$s\geqslant 0$，$X(\omega)$与$X_s(\omega)$有相同的分布.

(2) 对任意n元实值博雷尔可测函数$f(y_1,\cdots,y_n)$，有

$$E[f(x(t_1),\cdots,x(t_n))] = E[f(x(t_1+s),\cdots,x(t_n+s))]$$

(3) 对任意定义于 R_1^T 上的实值 $\mathscr{B}_1^T$ 可测函数 $\varphi(e(\cdot))$，有

$$E\{\varphi(X)\} = E\{\varphi(X_s)\}.$$

一般，强平稳过程未必有二阶矩，因而不一定是强平稳过程. 另一方面，弱平稳过程不一定是强平稳过程. 但对于正态过程而言，强平稳与弱平稳是相同的.

23.6 线性时不变系统

23.6.1 引言

在工业生产和科学实验中遇到的各类随机过程，往往是与系统联系的，即对各种"输入"按某种规则产生"输出". 而研究系统的各种特性就是讨论输入输出间的联系及各种指标间的联系. 例如，电路分析中，电路本身是一个系统，电路中电子热运动——热噪声就是输入，而电路输出端的电压或电流波动就是输出，习惯上系统用一个方框表示，如图 23.3 所示.

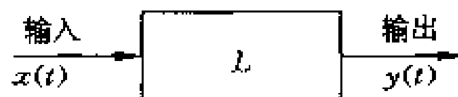


图 23.3

记 $x(t)$ 表示输入，$y(t)$ 表示输出，L 表示系统作用，即

$$y(t) = L[x(t)].$$

数学中 L 称为算子，它作用在 $x(t)$ 上的结果是 $y(t)$，L 由系统的物理特性决定，这里输入 $x(t)$ 可以是随机的.

23.6.2 线性时不变系统

定义 23.6.1 设系统用 L 表示，若

$$y_i(t) = L[x_i(t)], \quad i = 1,2,$$

对任意常数 α,β 都有

$$L[\alpha x_1(t) + \beta x_2(t)] = \alpha L[x_1(t)] + \beta L(x_2(t)]$$
$$= \alpha y_1(t) + \beta y_2(t),$$

则称系统 L 是**线性系统**(linear system).

如果系统 L 对任意$\tau>0$,都有

$$L(x(t+\tau)] = y(t+\tau),$$

则称系统 L 是**时不变的**(time invariant system).

由定义可知,线性时不变系统满足叠加原理,并且输出对输入的依赖关系不随时间的推移而改变. 一个物理系统处于稳态后,一般认为它是时不变的.

定义 23.6.2 设系统 L 满足

$$L[\lim_{n\to\infty} x_n(t)] = \lim_{n\to\infty} L[x_n(t)],$$

则称"L"具有连续性.

定理 23.6.3 设"L"为线性时不变系统,如果输入为 $x(t)=e^{i\omega t}$,则它的输出为

$$y(t) = H(\omega)e^{i\omega t},$$

其中 $H(\omega)$是一个复值函数,称为**频率响应函数**,简称**频率响应**(frequency response function).

该定理说明,当输入 $x(t)$是一个简谐波时,线性时不变系统的输出仍是简谐波,且频率不变. 若

$$H(\omega) = A(\omega)e^{i\theta(\omega)} \tag{23.15}$$

则 $A(\omega) = |H(\omega)|$ 称为系统的**振幅特性**(amplitude characteristic). $\theta(\omega)$是 $H(\omega)$的幅角,称为系统的**相位特性**(phase characteristic).

定理 23.6.4 设"L"是连续的时不变系统,具有频率响应 $H(\omega)$,当输入为 $x(t)$时,输出 $y(t)$可表示为

$$y(t) = \int_{-\infty}^{\infty} h(t-s)x(s)\mathrm{d}s,$$

其中

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}\omega. \qquad (23.16)$$

如果输入是一个**单位脉冲** $\delta(t)$(unit pulse),则有

$$y(t) = \int_{-\infty}^{\infty} h(t-s)\delta(s)\mathrm{d}s = h(t),$$

所以 $h(t)$ 可以看成系统在输入一个脉冲单位时的输出,故称它为**脉冲响应**(impluse response). 脉冲响应与频率响应构成一对傅里叶变换.

23.6.3 输出与输入是随机过程情况

定理 23.6.5 设 L 是线性时不变且对均方收敛连续的系统,则

(1) 当输入 $X(t)$ 是实平稳过程时,输出 $Y(t)$ 也是实平稳过程.

(2) $X(t)$ 与 $Y(t)$ 随机谱函数之间有

$$\zeta_Y(\omega) = \int_{-\infty}^{\omega} H(\mu)\mathrm{d}\zeta_X(\mu),$$

其中 $H(\omega)$ 是系统的频率响应.

(3) $X(t)$ 与 $Y(t)$ 的谱函数之间有

$$F_y(\omega) = \int_{-\infty}^{\omega} |H(\mu)|^2\mathrm{d}F_X(\mu).$$

(4) 如果 $X(t)$ 有谱密度,那么 $Y(t)$ 也有谱密度,且谱密度之间有

$$f_y(\omega) = |H(\omega)|^2 f_X(\omega).$$

23.6.4 平稳相关过程与互谱函数

定义 23.6.6 设 $X_T=\{X(t), -\infty<t<+\infty\}$, $Y_T=\{y(t),$

$-\infty<t<+\infty\}$均为平稳过程，若对任何 h，

$$E[X(t+h)\overline{Y(s+h)}]=E[X(t)\overline{Y(s)}]$$

恒成立，则称 X_T 与 Y_T **平稳相关**的(stationary correlative).

定义 23.6.7 当 $X(t),Y(t)$为两个平稳相关的平稳过程时，

$$B_{XY}(\tau)=E[X(t+\tau)\overline{Y(t)}],$$

称为 $X(t)$与 $Y(t)$的**互相关函数**(cross correlation function).

性质 23.6.8 互相关函数的性质

(1) $B_{XY}(\tau)=\overline{B_{YX}(-\tau)}$,

(2) $|B_{XY}(\tau)|^2\leqslant B_X(0)B_Y(0)$.

定义 23.6.9 若 $X(t),Y(t)$是平稳相关的，且

$$\sum_{\tau=-\infty}^{\infty}|B_{XY}(\tau)|<+\infty,\text{则}$$

$$f_{XY}(\lambda)=\frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}B_{XY}(\tau)e^{-i\tau\lambda},\quad -\pi\leqslant\lambda\leqslant\pi,\tag{23.17}$$

为 $X(t)$与 $Y(t)$的**互谱密度**(cross spectral density).

性质 23.6.10 互谱密度的性质

$$f_{XY}(\lambda)=\overline{f_{YX}(\lambda)}.$$

一般，$f_{XY}(\lambda)$都是复函数，没有非负的性质.

对 $B_{XY}(t)$有谱分解式为

$$B_{XY}(t)=\int_{-\infty}^{\infty}e^{i\lambda t}dE_{XY}(\lambda).\tag{23.18}$$

特别，当 $F_{XY}(\lambda)$绝对连续时，更可表示为

$$B_{XY}(t)=\int_{-\infty}^{\infty}e^{i\lambda t}f_{XY}(\lambda)d\lambda,\tag{23.19}$$

称 $F_{XY}(\lambda)$为 $X(t)$与 $Y(t)$的**互谱函数**(cross spectral function)，互谱函数是有界变差函数，不一定是单调不减函数.

定理 23.6.11 设 $X(t),Y(t)$是线性时不变，且对均方收敛具有连续性的系统输入和输出，它们的谱密度为 $f_X(\lambda),f_Y(\lambda)$. 则

$$f_Y(\lambda)=H(\lambda)f_X(\lambda),\tag{23.20}$$

其中 $H(\lambda)$是系统的频率响应.

23.7 平稳时间序列的线性模型

23.7.1 线性模型

设$\{x_t\}$为一零均值实平稳序列,它满足下列差分方程

$$x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \tag{23.21}$$

其中 ε_t 为正态平稳白噪声(例 23.1.6),$E(\varepsilon_t)=0, E(\varepsilon_t^2)=\sigma_\varepsilon^2$,且假定

$$E(x_t \varepsilon_s) = 0 \ (t < s), \tag{23.22}$$

$\varphi_1, \varphi_2, \cdots, \varphi_p$ 和 $\theta_1, \theta_2, \cdots, \theta_q$ 都是常数,p,q 为非负整数,用 B 表示**延迟算子**,或推移算子(delay operator, backward shift operator),$Bx_t = x_{t-1}, B^k x_t = x_{t-k}$,并令 $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$,$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$,则式(23.21)可缩写为

$$\varphi(B) x_t = \theta(B) \varepsilon_t. \tag{23.23}$$

满足上式的线性模型常考虑以下三种形式:AR(p),MA(q)和ARMA(p,q).

1. 自回归模型——AR(p)

定义 23.7.1 在(23.21)式中,若 $q=0$ 且满足条件:方程$\varphi(z)=0$的根皆在单位圆外,则称$\{x_t\}$满足 p 阶**平稳自回归模型**(stationary autoregressive model),记作 AR(p),其中 $\varphi(z) \triangleq 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$. 此时(23.22)成为 $\varphi(B) x_t = \varepsilon_t$. 称 p 为模型的阶,上述条件称为**平稳性条件**(stationarity condition).

2. 滑动平均模型——MA(q)

定义 23.7.2 在(23.21)式中,若 $p=0$,且满足条件:方程$\theta(z)=0$的根皆在单位圆外,则称$\{x_t\}$满足 q 阶**可逆滑动平均模型**

(invertible moving average model)，记作 MA(q). 此时(23.23)式成为 $x_t=\theta(B)\varepsilon_t$，称 q 为模型的阶，上述条件称为**可逆性条件**(invertibility condition).

3. 自回归滑动平均混合模型——ARMA(p,q)

定义 23.7.3 设$\{x_t\}$满足(23.21)式，$\varphi(z)$，$\theta(z)$分别满足平稳性，可逆性条件，且 $\varphi(z)$与 $\theta(z)$无公共因子，则称$\{x_t\}$满足**自回归一滑动平均混合模型**(mixed autoregressive-moving average model). 记作 ARMA(p,q)，(p,q)为模型的阶.

x_t 与 ε_t 相互关系的另外两种表达形式如下：传递形式——ARMA 模型平稳解的一种表达方式

$$x_t=\varphi^{-1}(B)\theta(B)\varepsilon_t \triangleq \psi(B)\varepsilon_t \triangleq \sum_{k=1}^{\infty}\psi_k B^k\varepsilon_t, \quad (23.24)$$

其中 $\psi(B)=\sum_{k=0}^{\infty}\psi_k B^k=\varphi^{-1}(B)\theta(B)$.

式(23.24)就是(23.23)的传递形式.

逆转形式——ARMA 模型中的白噪声 ε_t 表示成 x_t 的形式.

$$\varepsilon_t=\theta^{-1}(B)\varphi(B)x_t \triangleq \pi(B)x_t \triangleq x_t-\sum_{k=1}^{\infty}\pi_k B^k x_t \quad (23.25)$$

其中 $\pi(B)=1-\sum_{k=1}^{\infty}\pi_k B^k$.

式(23.25)就是(23.7.2)的逆转形式.

例 23.7.4 平稳性条件　AR(1)模型：形如 $x_t-\varphi_1 x_{t-1}=\varepsilon_t$ 相应的代数方程：$\varphi(z)=1-\varphi_1 z=0$，其根 $z_1=\frac{1}{\varphi_1}$. 欲使$|z_1|>1$，只需$|\varphi_1|<1$，此即平稳性条件.

23.7.2 ARMA 序列的自相关函数

1. MA(q)序列的自相关函数

自协方差函数为

$$
r_k \triangleq E x_t x_{t-k} = \begin{cases} \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2), & k=0, \\ \sigma_\varepsilon^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q), & 1 \leqslant k \leqslant q, \\ 0, & k>q. \end{cases}
$$

$r_k = r_{-k}$, $k<0$.

而自相关函数为

$$
\rho_k \triangleq \frac{r_k}{r_0} = \begin{cases} 1, & k=0, \\ \dfrac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2}, & 1 \leqslant k \leqslant q, \\ 0, & k>q. \end{cases} \tag{23.26}
$$

$$
\rho_k = \rho_{-k}, \quad k<0.
$$

MA(q)的自相关函数 ρ_k 在 $k>q$ 以后全部为零,此性质称为"截尾"性(cut off),这是MA(q)模型的特性.

例 23.7.5 自相关函数 MA(1)模型 $x_t=\varepsilon_t-\theta_1\varepsilon_{t-1}$,其自相关函数

$$
\rho_k = \begin{cases} 1, & k=0, \\ -\theta_1/(1+\theta_1^2), & k=\pm 1, \\ 0, & \text{其它}. \end{cases}
$$

2. AR(p)序列的自相关函数

AR(p)序列的自相关函数满足差分方程

$$
\rho_k = \varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}+\cdots+\varphi_p\rho_{k-p}, \quad k>0. \tag{23.27}
$$

即 $\varphi(B)\rho_k=0$,根据差分方程理论知,当 $\varphi(B)=0$ 无重根时,有

$$
\rho_k = c_1 z_1^{-k}+c_2 z_2^{-k}+\cdots+c_p z_p^{-k} \quad (k>-p),
$$

其中 $z_1,z_2,\cdots,z_p$ 是 $\varphi(B)=0$ 的根,$c_1,c_2,\cdots,c_p$ 是待定系数,它们可由下列方程组解出.方程组是根据 $\rho_k=\rho_{-k}$,分别取 $k=0,1,2,\cdots,p-1$ 而得.

$$\begin{cases} \sum_{i=1}^{p} c_i = 1, \\ \sum_{i=1}^{p} c_i(z_i^k - z_i^{-k}) = 0, \quad k = 1,2,\cdots, p = 1. \end{cases}$$

当 $\varphi(B)=0$ 有重根，设前 r 个根相同. 即 $z_1=z_2=\cdots=z_r$，则

$$\rho_k = (c_1 + c_2 k + \cdots + c_r k^{r-1}) z_1^{-k} + c_{r+1} z_{r+1}^{-k} + \cdots + c_p z_p^{-k}.$$

利用 ρ_k 的对称性，可求出诸系数 c_i.

无论 $\varphi(B)=0$ 有无重根，ρ_k 都不是截尾的序列，但 $|\rho_k|$ 又被负指数函数所控制. 这种 ρ_k 的尾部不全为零，但又被负指数函数控制的性质称为**"拖尾"**性(dies cut gradually). 这是 AR(p)序列的特性.

例 23.7.6 AR 序列的拖尾性 AR(2)序列 $x_t-\varphi_1 x_{t-1}-\varphi_2 x_{t-2}=\epsilon_t$，其自相关函数满足方程 $\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}, k>0$. 当 $\varphi(z)=1-\varphi_1 z-\varphi_2 z^2=0$ 的根 $z_1\neq z_2$ 时，其解 ρ_k 为 $\rho_k=c_1 z_1^{-k}+c_2 z_2^{-k}, k\geqslant -1$.

由
$$\begin{cases} c_1+c_2=1, \\ c_1(z_1-z_1^{-1})+c_2(z_2-z_2^{-1})=0, \end{cases}$$

解出 $c_1=\dfrac{z_2-z_2^{-1}}{(z_2-z_2^{-1})-(z_1-z_1^{-1})}$，　$c_2=\dfrac{z_1-z_1^{-1}}{(z_1-z_1^{-1})-(z_2-z_2^{-1})}$.

当 $z_1=z_2$ 时，$\rho_k=(c_1+c_2 k)z_1^{-k}(k\geqslant -1)$，由 $\rho_0=1$，可知 $c_1=1$；又由 $\rho_1=\rho_{-1}$，得 $c_2=\dfrac{z_1^2-1}{1+z_1^2}$. 由于 $|z_1|>1$，$|z_2|>1$，因此 ρ_k 总被负指数函数所控制.

3. ARMA(p,q)的自相关函数

对于 ARMA(p,q)模型，利用(23.24)形式，先求出 x_t 与 ϵ_t 的互协方差系数.

$$r_k(x,\epsilon) \triangleq E(x_t \epsilon_{t+k}) = \begin{cases} \sigma_\epsilon^2 \psi_{-k}, & k \leqslant 0, \\ 0, & k > 0. \end{cases} \tag{23.28}$$

把(23.21)式两边同乘以 x_{t-k}，再求均值，得到

$$r_k - \varphi_1 r_{k-1} - \cdots - \varphi_p r_{k-p}$$

$$= \begin{cases} r_k(x,\varepsilon) - \theta_1 r_{k-1}(x,\varepsilon) - \cdots - \theta_q r_{k-q}(x,\varepsilon), & k \leqslant q, \\ 0 & k > q. \end{cases} \tag{23.29}$$

故当 $k>q$ 时，有 $\varphi(B)\rho_k=0$. 这是类似于式(23.27)的方程，关于解的形式已讨论过了. 不同之处在于确定初始值的方法. 如当 $\varphi(z)=0$ 无重根时，知道了初始值 $\rho_q,\rho_{q-1},\cdots,\rho_{q-p+1}$，通解就可写为

$$\rho_k = \sum_{i=1}^{p} c_i z_i^{-k}, \ k>q-p.$$

例 23.7.7 ARMA 模型的自相关函数　求 ARMA(1,1)模型 $x_t-0.3x_{t-1}=\varepsilon_t-0.6\varepsilon_{t-1}$ 的自相关函数.

$\varphi_1=0.3<1$，$\theta_1=0.6<1$，故满足平稳性、可逆性条件.

$\varphi(z)=1-0.3z=0$，得 $z_1=\dfrac{10}{3}$. 又 $p=q=1$ 故对 $k>0$，有

$$\rho_k=c_1 z_1^{-k}=c_1\left(\frac{10}{3}\right)^{-k}=c_1(0.3)^k.$$

由(例 23.7.6)，$r_0(x,\varepsilon)=\sigma_\varepsilon^2\varphi_0=\sigma_\varepsilon^2$(因 $\varphi_0=1$)，

$$r_{-1}(x,\varepsilon)=\sigma_\varepsilon^2\varphi_1. \quad r_1(x,\varepsilon)=0.$$

而由式(23.24)，可得 $\varphi_1=\varphi_1-\theta_1=-0.3$，由式(23.29)令 $k=0,1$，得

$$\begin{cases} r_0 - 0.3r_{-1} = r_0(x,\varepsilon) - 0.6r_{-1}(x,\varepsilon) = 1.18\sigma_\varepsilon^2, \\ r_1 - 0.3r_0 = r_1(x,\varepsilon) - 0.6r_0(x,\varepsilon) = -0.6\sigma_\varepsilon^2. \end{cases}$$

又 $r_1=r_{-1}$，解得 $r_0=\dfrac{1}{0.91}\sigma_\varepsilon^2$，$r_1=\dfrac{-246}{910}\sigma_\varepsilon^2$.

故 $\rho_1=-0.246$，$c_1=-0.82$. 所以

$$\rho_k=-0.82(0.3)^k, \quad k>0.$$

23.7.3　ARMA 序列的偏相关函数

设$\{x_t\}$为零均值实平稳列，用 $x_{t-1},x_{t-2},\cdots,x_{t-k}$对 x_t 作最小

方差估计,即选择系数 $\varphi_{k1},\varphi_{k2},\cdots,\varphi_{kk}$,使得

$$\delta \triangleq E\left(x_t - \sum_{j=1}^{k}\varphi_{kj}x_{t-j}\right)^2 \tag{23.30}$$

达到极小值.为此,分别对 $\varphi_{kj}(j=1,2,\cdots,k)$ 求 δ 的偏导数 $\frac{\partial \delta}{\partial \varphi_{kj}}$,并令其为零,得到 φ_{kj} 应满足的方程组

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}. \tag{23.31}$$

定义 23.7.8 设 $\varphi_{kj},j=1,2,\cdots,k$ 使式(23.30)达到极小值,贝称序列 $\{\varphi_{kk}\}(k\geqslant 1)$ 为 $\{x_t\}$ 的**偏相关函数**(partial correlation function).方程组(23.31)称为**耶鲁-沃克**(Yule-Walker)方程.等式左边的矩阵称为**特普利茨**(Toeplitz)**矩阵**.

φ_{kk} 的概率意义是:在给定 $x_{t-1},x_{t-2},\cdots,x_{t-k+1}$ 的条件下,x_t 与 x_{t-k} 的条件相关系数,即 $\varphi_{kk}=(E\tilde{X}_t\tilde{X}_{t-k})(E\tilde{X}_t^2\cdot E\tilde{X}_{t-k}^2)^{-\frac{1}{2}}$,其中 $\tilde{X}_t=x_t-E(x_t|x_{t-1},\cdots,x_{t-k+1})$.$\varphi_{kk}$ 也是被负指数函数所控制的.

1. AR(p)的偏相关函数

将 AR(p)模型 $x_t=\sum_{j=1}^{p}\varphi_j x_{t-j}+\epsilon_t$ 代入(23.30),整理得

$$\delta = E(\epsilon_t^2) + 2E\left(\epsilon_t\left(\sum_{j=1}^{p}(\varphi_j-\varphi_{kj})x_{t-j} - \sum_{j=p+1}^{k}\varphi_{kj}x_{t-j}\right)\right) + E\left(\left(\sum_{j=1}^{p}(\varphi_j-\varphi_{kj})x_{t-j} - \sum_{j=p+1}^{k}\varphi_{kj}x_{t-j}\right)^2\right),$$

欲使 δ 达到最小值,应使第三项为零,故应使

$$\begin{cases} \varphi_{kj}=\varphi_j, & 1\leqslant j\leqslant p, \\ \varphi_{kj}=0, & j=p+1,p+2,\cdots,k. \end{cases} \tag{23.32}$$

故当 $k>p$ 后，必有 $\varphi_{kk}=0$，这说明 AR(p)的偏相关函数是“截尾”的，而且“截尾”处的 k 值就是模型的阶数.

2. ARMA(p,q)和 MA(q)的偏相关函数

由偏相关函数定义，为使(23.30)中的 δ 达到最小值，需解方程组(23.31)，即可求出偏相关函数 φ_{kk} 的值.

算法 23.7.9 偏相关函数的递推算法

公式如下：

$$\begin{cases}\varphi_{11}=\rho_1,\\ \varphi_{k+1,k+1}=\left(\rho_{k+1}-\sum\limits_{j=1}^{k}\rho_{k+1-j}\varphi_{kj}\right)\left(1-\sum\limits_{j=1}^{k}\rho_j\varphi_{kj}\right)^{-1},\\ \varphi_{k+1,j}=\varphi_{kj}-\varphi_{k+1,k+1}\varphi_{k,k-j+1}, \quad j=1,2,\cdots,k.\end{cases}$$

利用此公式比直接用式(23.31)求解计算量要小.

23.7.4 ARMA 模型的谱密度函数

定理 23.7.10 设 $\{x_t\}$ 满足差分方程(23.21)，其中 $\{\varepsilon_t\}$ 是一白噪声序列，

$$E(\varepsilon_t)=0, \quad E(\varepsilon_t\varepsilon_s)=\begin{cases}\sigma_\varepsilon^2, & t=s,\\ 0, & t\neq s.\end{cases}$$

$\varphi(B)$ 与 $\theta(B)$ 同前. 设 $f_x(\omega)$ 是 x_t 的谱密度，则

$$f_x(\omega)=\frac{\sigma_\varepsilon^2}{2\pi}\frac{|\theta(\mathrm{e}^{-\mathrm{i}\omega})|^2}{|\varphi(\mathrm{e}^{-\mathrm{i}\omega})|^2}, \quad -\pi\leqslant\omega\leqslant\pi. \qquad (23.33)$$

系 23.7.11 设 $\{x_t\}$ 满足 $\varphi(B)x_t=\varepsilon_t$，则其谱密度为

$$f_x(\omega)=\frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{|\varphi(\mathrm{e}^{-\mathrm{i}\omega})|^2}, \quad -\pi\leqslant\omega\leqslant\pi.$$

系 23.7.12 设 $\{x_t\}$ 满足 $\theta(B)\varepsilon_t=x_t$，则其谱密度为 $f_x(\omega)=\dfrac{\sigma_\varepsilon^2}{2\pi}|\theta(\mathrm{e}^{-\mathrm{i}\omega})|^2, -\pi\leqslant\omega\leqslant\pi$.

表 23.2　ARMA 序列的分类性质一览表

类别＼模型	AR(p)	MA(q)	ARMA(p,q)
模型方程	$\varphi(B)x_t=\varepsilon_t$	$x_t=\theta(B)\varepsilon_t$	$\varphi(B)x_t=\theta(B)\varepsilon_t$
平稳条件	$\varphi(B)=0$ 的根的模大于 1	无条件	$\varphi(B)=0$ 的根的模>1
可逆条件	无条件	$\theta(B)=0$ 的根的模大于 1	$\theta(B)=0$ 的根的模>1
传递形式	$x_t=\varphi^{-1}(B)\varepsilon_t$	$x_t=\theta(B)\varepsilon_t$	$x_t=\varphi^{-1}(B)\theta(B)\varepsilon_t$ $=\sum_{k=0}^{\infty}\varphi_k\varepsilon_{t-k}$
逆转形式	$\varepsilon_t=\varphi(B)x_t$	$\varepsilon_t=\theta^{-1}(B)x_t$	$\varepsilon_t=\theta^{-1}(B)\varphi(B)x_t$ $=x_t-\sum_{j=1}^{\infty}\pi_j x_{t-j}$
自相关函数	拖尾	截尾	拖尾
偏相关函数	截尾	拖尾	拖尾

23.8　模型的识别、参数估计与检验

本节将给出如何由样本值来识别模型并估计模型的参数及进行模型的检验.

23.8.1　自相关与偏相关函数的估计

1. 均值估计 $\hat{\mu}$

设 $x_1,x_2,\cdots,x_N$ 是平稳正态序列 $\{x_t\}$ 的一段样本值，N 为样本长度. 设 $Ex_t=\mu$. 则样本均值 $\bar{x}$ 作为 μ 的估计，即

$$\hat{\mu} = \bar{x} \triangleq \frac{1}{N}\sum_{t=1}^{N} x_t. \tag{23.34}$$

2. 自协方差函数的估计 $\hat{r}_k$

为方便计，设$\{x_t\}$为平稳正态零均值序列. 否则，可令 $w_t = x_t - \bar{x}$ 即可得**自协方差函数的估计**(estimation of autocovarrance function)

$$\hat{r}_k = \hat{r}_{-k} \triangleq \frac{1}{N}\sum_{t=1}^{N-k} x_t x_{t+k}, \quad k = 0,1,2,\cdots,M, \tag{23.35}$$

$$\hat{\sigma}^2 = \hat{r}_0.$$

此估计保证非负定性，但只是渐近无偏的. 一般取 $N \geqslant 50$，常用 $M = N/10$左右，不超过 $N/4$.

3. 自相关函数估计 $\hat{\rho}_k$

自相关函数估计(estimation of autocorrelation functions)为

$$\hat{\rho}_k = \hat{\rho}_{-k} = \frac{\hat{r}_k}{\hat{r}_0}, \quad k = 0,1,2,\cdots,M. \tag{23.36}$$

或采用无偏估计，但不保证非负定性. 即取

$$\hat{r}_k^* = \hat{r}_{-k}^* = \frac{1}{N-k}\sum_{t=1}^{N-k} x_t x_{t+k}, \quad k = 0,1,2,\cdots,M. \tag{23.37}$$

$$\hat{\rho}_k^* = \hat{\rho}_{-k}^* = \hat{r}_k^* / \hat{r}_0^*.$$

4. 偏相关函数估计 $\hat{\varphi}_{kk}$

解方程组

$$\begin{pmatrix} \hat{\varphi}_{k1} \\ \hat{\varphi}_{k2} \\ \vdots \\ \hat{\varphi}_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \vdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_k \end{pmatrix}, \tag{23.38}$$

即可得到 $\varphi_{kk}(k=1,2,\cdots)$**偏相关函数的估计**(estimation of partial correlation function)或用递推算法 23.7.9 也可得到 $\hat{\varphi}_{kk}$.

23.8.2 模型的初步识别

根据 ARMA 序列自相关及偏相关函数的截尾性质，可初步判别$\{x_t\}$属于哪类模型，即初步的**模型识别**(model recognition, discriminate of the model).

1. $\hat{\rho}_k$ 截尾性的判断

若$\{x_t\}$是一个 MA(q)序列，则当 N 充分大时，$\hat{\rho}_k$ 渐近正态分布，即

$$\hat{\rho}_k \sim N\left(0, \frac{1}{N}\left(1+2\sum_{k=1}^{q}\hat{\rho}_k^2\right)\right),$$

当 q 不很大而 N 很大时，$\frac{1}{N}\left(1+2\sum_{k=1}^{q}\hat{\rho}_k^2\right) \approx \frac{1}{N}$.

故又可近似为

$$\hat{\rho}_k \sim N(0, 1/N. \tag{23.39}$$

一般对某个选定的 q，检查 $\hat{\rho}_k$，$k=q+1, q+2, \cdots, M$. 若满足 $|\hat{\rho}_k| < 2/\sqrt{N}$的个数在 95%以上，则可认为 $\hat{\rho}_k$ 是 q 步截尾的，q 就是模型的阶数，q 可从小到大逐个进行检验确定.

2. $\hat{\varphi}_{kk}$截尾性的判断

若$\{x_t\}$是一个 AR(p)序列，当 N 充分大时，也有 $\hat{\varphi}_{kk}$渐近于正态分布，即

$$\hat{\varphi}_{kk} \sim N(0, 1/N),$$

故仍可用 1 中的方法来判断 $\hat{\varphi}_{kk}$的"截尾"性，并选定 p 的值.

3. ARMA，(p,q)模型的定阶

若$\{x_t\}$的 $\hat{\rho}_k$ 和 $\hat{\varphi}_{kk}$ 都不截尾，又被负指数型的数列所控制，则应判断为 ARMA 序列，但尚不能确定阶数. 而由于实际应用中的(p,q)一般都比较低，因此可采用从低阶到高阶逐个取(p,q)为(1,1)，(1,2)，(2,1)，…等值尝试的办法，即先认定(p,q)为某值，然后进行参数估计并定出模型，然后检验此模型是否被接受，若不

被接受就调整(p,q)值，再重新做参数估计和模型检验，直到被接受为止.

23.8.3 模型参数的矩估计法

利用各种模型中 r_k 或 ρ_k 与模型参数的关系，从估计 $\hat{r}_k,\hat{\rho}_k$ 中解出参数的估计值，此称为模型参数的**矩估计法**(method of moments).

1. AR(p)模型的参数矩估计

AR(p)模型为 $x_t-\varphi_1 x_{t-1}-\cdots-\varphi_p x_{t-p}=\varepsilon_t$，待估计参数为 φ_1，$\varphi_2,\cdots,\varphi_p$ 和 σ_ε^2，$\{\varphi_j\}$的 Yule-Walker 估计为

$$\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{pmatrix}=\begin{pmatrix}1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1}\\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2}\\ \cdots & \cdots & \cdots & \cdots\\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \vdots & 1\end{pmatrix}^{-1}\begin{pmatrix}\hat{\rho}_1\\ \hat{\rho}_2\\ \vdots\\ \hat{\rho}_p\end{pmatrix}, \tag{23.40}$$

$$\hat{\sigma}_\varepsilon^2=\hat{r}_0-\sum_{j=1}^{p}\hat{\varphi}_j\hat{r}_j.$$

这里 $\hat{\rho}_j,\hat{r}_j$ 由(23.35)和(23.36)式根据样本值算出.

2. MA(q)模型的参数矩估计

MA(q)模型为 $x_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$，

待估计参数为 $\theta_1,\theta_2,\cdots,\theta_q$ 和 σ_ε^2.

由矩估计式：

$$\hat{r}_k=\begin{cases}\hat{\sigma}_\varepsilon^2(1+\hat{\theta}_1^2+\cdots+\hat{\theta}_q^2), & k=0,\\ \hat{\sigma}_\varepsilon^2(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q), & 1\leqslant k\leqslant q,\\ 0, & k>q.\end{cases} \tag{23.41}$$

解出 $\hat{\sigma}_\varepsilon^2,\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_q$ 即可.不难得到如下关系：

$$\hat{\sigma}_\varepsilon^2=\hat{r}_0\Big(\sum_{i=1}^{q}\hat{\theta}_i^2\Big)^{-1},\quad \theta_0=1.$$

$$\hat{\theta}_k = \frac{-\hat{r}_k}{\hat{\sigma}_\varepsilon^2} + (\hat{\theta}_1\hat{\theta}_{k+1} + \cdots + \hat{\theta}_{q-k}\hat{\theta}_q), \quad 1 \leqslant k \leqslant q.$$

上述 $\hat{r}_k$ 由(23.35)式算出.

例 23.8.1 求 MA(1)模型的参数估计 由(23.41)及 $q=1$,有

$$\begin{cases} \hat{r}_0 = \hat{\sigma}_\varepsilon^2(1 + \hat{\theta}_1^2), \\ \hat{r}_1 = -\hat{\theta}_1\hat{\theta}_\varepsilon^2. \end{cases}$$

由第二式得 $\hat{\theta}_1 = -\hat{r}_1/\hat{\sigma}_\varepsilon^2$,代入第一式化简得$(\hat{\sigma}_\varepsilon^2)^2 - \hat{r}_0\hat{\sigma}_\varepsilon^2 + \hat{r}_1^2 = 0$,解方程得 $\hat{\sigma}_\varepsilon^2 = \frac{1}{2}(\hat{r}_0 \pm \sqrt{\hat{r}_0^2 - 4\hat{r}_1^2})$,故 $\hat{\theta}_1 = -2\hat{r}_1/(\hat{r}_0 \pm \sqrt{\hat{r}_0^2 - 4\hat{r}_1^2})$. 由可逆性条件,需$|\theta_1| < 1$,故 $\hat{\theta}_1$ 中的分母应取正号,即有

$$\hat{\theta}_1 = -2\hat{r}_1 \Big/ \left(\hat{r}_0 + \sqrt{\hat{r}_0^2 - 4\hat{r}_1^2}\right),$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2}\left(\hat{r}_0 + \sqrt{\hat{r}_0^2 - 4\hat{r}_1^2}\right).$$

3. ARMA(p,q)模型的参数矩估计

ARMA(p,q)模型为 $x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$待估参数为 $\varphi_1, \varphi_2, \cdots, \varphi_p, \theta_1, \theta_2, \cdots, \theta_q$ 和 σ_ε^2.

首先给出自回归参数 φ_i 的矩估计.

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} \hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p-1} \\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p+2} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \vdots & \hat{\rho}_q \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_{q+1} \\ \hat{\rho}_{q+2} \\ \vdots \\ \hat{\rho}_{q+p} \end{pmatrix}.$$

再令 $\eta_t = x_t - \sum_{j=1}^{p} \hat{\varphi}_j x_{t-j}$,其协方差函数为

$$r_\eta(k) = E(\eta_t\eta_{t+k}) = \sum_{i,j=0}^{p} \hat{\varphi}_j\hat{\varphi}_i r_{k+j-i}, \quad \text{记 } \hat{\varphi}_0 = -1.$$

$$\hat{r}_\eta(k) = \sum_{i,j=0}^{p} \hat{\varphi}_i\hat{\varphi}_j\hat{r}_{k+j-i}$$

用 η_t 近似(23.21)式的左方,则可有 MA(q)序列 $\eta_t=\theta(B)\varepsilon_t$,用 2 中的方法求出 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_q$ 和 $\hat{\sigma}_\varepsilon^2$,从而就求出了 ARMA($p,q$)模型的全部参数的矩估计.

23.8.4 模型参数的最小二乘估计

1. AR(p)模型参数的最小二乘估计

AR(p)模型为 $x_t=\sum\limits_{j=1}^{p}\varphi_j x_{t-j}+\varepsilon_t$, 令 $\boldsymbol{\varphi}=(\varphi_1,\varphi_2,\cdots,\varphi_p)^{\mathrm{T}}$,则 $\boldsymbol{\varphi}$ 的最小二乘估计 $\hat{\boldsymbol{\varphi}}^L$ 应使残差平方和 $S(\boldsymbol{\varphi})$达到最小,这里

$$S(\boldsymbol{\varphi})=\sum_{t=p+1}^{N}\varepsilon_t^2=\sum_{t=p+1}^{N}\left(x_t-\sum_{j=1}^{p}\varphi_j x_{t-j}\right)^2,$$

取$\dfrac{\partial}{\partial\boldsymbol{\varphi}}S(\boldsymbol{\varphi})$,并令其为零,则当 N 较大时,可得 Ψ 的最小二乘估计的近似解为

$$\hat{\boldsymbol{\varphi}}^L=\begin{pmatrix}\hat{\varphi}_1^L\\\hat{\varphi}_2^L\\\vdots\\\hat{\varphi}_p^L\end{pmatrix}\doteq\begin{pmatrix}\hat{r}_0^L & \hat{r}_1^L & \cdots & \hat{r}_{p-1}^L\\\hat{r}_1^L & \hat{r}_0^L & \cdots & \hat{r}_{p-2}^L\\\cdots & \cdots & \cdots & \cdots\\\hat{r}_{p-1}^L & \hat{r}_{p-2}^L & \vdots & \hat{r}_0^L\end{pmatrix}^{-1}\begin{pmatrix}\hat{r}_1^L\\\hat{r}_2^L\\\vdots\\\hat{r}_p^L\end{pmatrix}. \tag{23.42}$$

其中

$$\hat{r}_{j-i}^L=\frac{1}{N}\sum_{t=p+1}^{N}x_{t-j}x_{t-i}. \tag{23.43}$$

令 $\hat{\varepsilon}_t=x_t-\sum\limits_{j=1}^{p}\hat{\varphi}_j x_{t-j}$, $t=p+1,p+2,\cdots,N$ 则

$$\hat{\sigma}_\varepsilon^2=\frac{1}{N-p}\sum_{t=p+1}^{N}\hat{\varepsilon}_t^2=\hat{r}_0^L-\hat{\varphi}_1^L\hat{r}_1^L-\cdots-\hat{\varphi}_p^L\hat{r}_p^L. \tag{23.44}$$

2. MA(q)与 ARMA(p,q)模型参数的最小二乘估计

ARMA(p,q)模型为

$$x_t=\sum_{j=1}^{p}\varphi_j x_{t-j}-\sum_{i=1}^{q}\theta_i\varepsilon_{t-i}+\varepsilon_t.$$

引入函数 $f_t(x_t,\mathscr{B})=\sum\limits_{j=1}^{p}\varphi_j x_{t-j}-\sum\limits_{i=1}^{q}\theta_i\varepsilon_{t-i}$,

其中 $x_t=(x_{t-1},x_{t-2},\cdots,x_{t-p})^T$，$\mathscr{B}=(\varphi_1,\cdots,\varphi_p,\theta_1,\cdots,\theta_q)^T$.

由 ARMA(p,q)模型的逆转形式有

$$\varepsilon_t = x_t - \sum_{j=1}^{p}\varphi_j x_{t-j} + \sum_{i=1}^{q}\theta_i\sum_{k=1}^{\infty}\pi_k x_{t-i-k}. \qquad (23.45)$$

由于只有 N 个样本 $x_1,x_2,\cdots,x_N$，故取当 $k\leqslant 0$ 时，$x_k=0$，这样模型可写成 $x_t=f_t(x_t,\mathscr{B})+\varepsilon_t$，$p+1\leqslant t\leqslant N$，利用递推算法可求出 $\hat{\varepsilon}_t$：

$$\hat{\varepsilon}_t = x_t - \sum_{j=1}^{p}\varphi_j x_{t-j} + \sum_{i=1}^{q}\theta_i\hat{\varepsilon}_{t-i} \qquad t = p+1,\cdots,N. \qquad (23.46)$$

其中规定：$\hat{\varepsilon}_p=\hat{\varepsilon}_{p-1}=\hat{\varepsilon}_{p-2}=\cdots=\hat{\varepsilon}_{p+1-q}=0$，残差平方和为

$$S(\mathscr{B}) = \sum_{t=p+1}^{N}\hat{\varepsilon}_t^2.$$

对每一组具体参数 $\mathscr{B}$，便可算出 $\hat{\varepsilon}_t(t=1,2,\cdots,N)$和 $S(\mathscr{B})$，从中选出最小的一个 $S(\hat{\mathscr{B}})$相应的参数 $\hat{\mathscr{B}}$，即为所求的最小二乘估计 $\hat{\mathscr{B}}^L$.

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N-(p+q)}\sum_{t=p+1}^{N}\hat{\varepsilon}_t^2.$$

例 23.8.2 求 MA(1)模型参数的最小二乘估计 因 $p=0$，所以规定 $\hat{\varepsilon}_0=0$，由式(23.46)递推得

$$\hat{\varepsilon}_1 = x_1 + \theta_1\hat{\varepsilon}_0 = x_1,$$

$$\hat{\varepsilon}_2 = x_2 + \theta_1\hat{\varepsilon}_1 = x_2 + \theta_1 x_1,$$

$$\cdots\cdots,$$

$$\hat{\varepsilon}_k = x_k + \theta_1\hat{\varepsilon}_{k-1} = \sum_{j=0}^{k-1}\theta_1^j x_{t-j}.$$

于是

$$S(\theta_1) = \sum_{k=1}^{N}\hat{\varepsilon}_k^2 = \sum_{k=1}^{N}\Big(\sum_{j=0}^{k-1}\theta_1^j x_{k-j}\Big)^2.$$

要求使 $S(\theta_1)$达最小值的 θ_1^L 作为 θ_1 的最小二乘估计.

求 $S(\mathscr{B})$的最小值问题可采用动态规划中的各种数值解法或其它近似方法.除模型参数的上述两种估计外,还有最小平方和估计和极大似然估计等.

23.8.5 模型阶的确定

节 23.8.2 中给出了 MA(q),AR(p)和 ARMA(p,q)模型阶的判定方法,此外,模型拟合的**最小信息准则**(minimum information criterion)也称 **Akaike 信息准则**(简记为 AIC),也有广泛的应用.下面给出 ARMA(p,q)序列的 AIC 准则.此准则能在模型参数极大似然估计的基础上,对于 ARMA(p,q)模型的阶和相应参数同时给出一种最佳估计.

设$\{x_t\}$为平稳正态零均值的 ARMA(p,q)序列,对于给定样本 $\boldsymbol{X}_N=(x_1,x_2,\cdots,x_N)^{\mathrm{T}}$,当 N 充分大时,ARMA(p,q)模型的对数似然函数近似地可表示为

$$L_N(\hat{\mathscr{B}})\cong-\frac{N}{2}\ln 2\pi-\frac{N}{2}\ln\hat{\sigma}_\varepsilon^2-\frac{S(\hat{\mathscr{B}})}{2\hat{\sigma}_\varepsilon^2},\qquad(23.47)$$

其中 $\hat{\mathscr{B}}=(\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p,\hat{\theta}_1,\hat{\sigma}_2,\cdots,\hat{\theta}_q,\hat{\sigma}_\varepsilon^2)$是模型参数 $\mathscr{B}$ 的近似极大似然估计,往往是满足$\frac{\partial}{\partial\mathscr{B}}S(\mathscr{B})=0$ 的解.而此时 $\hat{\sigma}_\varepsilon^2=\frac{1}{N}S(\hat{\mathscr{B}})$,其中 $S(\mathscr{B})$是平方和函数,可证 $S(\hat{\mathscr{B}})=\sum\limits_{k=-\infty}^{N}\hat{\varepsilon}_k^2$.

$$\hat{\varepsilon}_k\simeq x_k-\hat{\varphi}_1x_{k-1}-\cdots-\hat{\varphi}_px_{k-p}+\hat{\theta}_1\hat{\varepsilon}_{k-1}+\cdots+\hat{\theta}_q\hat{\varepsilon}_{k-q},\qquad(23.48)$$

且约定 $\hat{\varepsilon}_k=0,k\leqslant q$,实际计算时取

$$\hat{\sigma}_\varepsilon^2\simeq\frac{1}{N-M}\sum_{k=M+1}^{N}\hat{\varepsilon}_k^2,\qquad(23.49)$$

式中 $\hat{\varepsilon}_k$ 由(23.48)递推,M 的大小依模型和数据的具体情况而定,一般取 $M=5q\sim10q$ 为宜.

在 N 充分大且固定的前提下，ARMA(p,q)模型拟合的最小信息准则应使

$$\mathrm{AIC} \triangleq N\ln\hat{\sigma}_\varepsilon^2 + 2(p+q+1) \tag{23.50}$$

达到最小，与之相应的阶数$(\hat{p},\hat{q})$及模型参数的极大似然估计就是在 AIC 准则下得到的理想阶数和参数.

当 $q=0$ 时，变为 AR(p)模型，此时利用

$$\mathrm{AIC}(p) = N\ln\hat{\sigma}_\varepsilon^2 + 2(p+1) \tag{23.51}$$

去找 p 值.

应用此准则所定的模型阶数并不是相容估计，且具体运用时需人为地给定阶数的最大范围，一般取 N，$\frac{N}{10}$，$\ln N$ 均可，目前 AIC 准则被广泛应用.

23.8.6 模型检验——残差的自相关函数法

设初步选定模型为 ARMA(p,q)，将估计出的参数和原始数据代入可得

$$\hat{\varepsilon}_t = x_t - \hat{\varphi}_1 x_{t-1} - \cdots - \hat{\varphi}_p x_{t-p} + \hat{\theta}_1 \varepsilon_{t-1} + \cdots + \hat{\theta}_q \varepsilon_{t-q}, \quad t = 1,2,\cdots,N$$

规定 $t\leqslant 0$ 时，$x_t=0$，$\varepsilon_t=0$，于是得残差的样本序列 $\hat{\varepsilon}_1,\hat{\varepsilon}_2,\cdots,\hat{\varepsilon}_N$，令

$$\hat{r}_{\varepsilon k} \triangleq \frac{1}{N}\sum_{t=1}^{N-k}\hat{\varepsilon}_t\hat{\varepsilon}_{t+k}, \quad k = 0,1,2,\cdots,M \quad \left(M = \frac{N}{10}\right),$$

$$\hat{\rho}_{\varepsilon k} = \hat{r}_{\varepsilon k}/\hat{r}_{\varepsilon 0},$$

$$Q_M = \sum_{k=1}^{M}(\sqrt{N}\hat{\rho}_k)^2. \tag{23.52}$$

若$\{\hat{\varepsilon}_t\}$是白噪声序列，则 $Q_M \sim \chi_M^2$ 即 Q_M 服从自由度为 M 的 χ^2 分布，检验法如下：原假设 $H_0: Q_M\sim\chi_M^2$，给定显著水平 α，可查出 $\chi_M^2(1-\alpha)$，对给定的样本值计算出 Q_M 值，若 $Q_M\leqslant\chi_M^2(1-\alpha)$，认为所选模型合适，若 $Q_M>\chi_M^2(1-\alpha)$，则认为此模型不合适.

23.9 ARMA 模型的谱分析

23.9.1 ARMA(p,q)模型谱估计的参数方法

在求出模型的参数估计值后，就可以得到功率谱密度的估计$\hat{f}_x(\omega)$：

例 23.9.1 AR(p)模型：$x_t - \sum_{j=1}^{p}\varphi_j x_{t-j} = \varepsilon_t$.

$$\hat{f}_x(\omega) = \frac{\hat{\sigma}_\varepsilon^2}{2\pi}\frac{1}{\left|1-\sum_{j=1}^{p}\hat{\varphi}_j e^{-ij\omega}\right|^2}, \quad -\pi \leqslant \omega \leqslant \pi.$$

例 23.9.2 MA(q)模型：$x_t = \varepsilon_t - \sum_{k=1}^{q}\theta_k\varepsilon_{t-k}$.

$$\hat{f}_x(\omega) = \frac{\hat{\sigma}_\varepsilon^2}{2\pi}\left|1-\sum_{k=1}^{q}\hat{\theta}_k e^{-ik\omega}\right|^2, \quad -\pi \leqslant \omega \leqslant \pi.$$

例 23.9.3 ARMA(p,q)模型：

$$x_t - \sum_{j=1}^{p}\varphi_j x_{t-j} = \varepsilon_t - \sum_{k=1}^{q}\theta_k\varepsilon_{t-k}.$$

$$\hat{f}_x(\omega) = \frac{\hat{\sigma}_\varepsilon^2}{2\pi}\left|\frac{1-\sum_{k=1}^{q}\hat{\theta}_k e^{-ik\omega}}{1-\sum_{j=1}^{p}\hat{\varphi}_j e^{-ij\omega}}\right|^2, \quad -\pi \leqslant \omega \leqslant \pi.$$

23.9.2 谱估计的非参数方法

1. 周期图的谱估计

定义 23.9.4 设$\{x_t\}$是零均值实平稳列，给定 N 个样本值 $x_1, x_2, \cdots, x_N$，在 $\omega\in[-\pi,\pi]$上，函数

$$I_N(\omega) = \frac{1}{2\pi}\sum_{k=1-N}^{N-1}\hat{r}_k e^{-ik\omega} = \frac{1}{2\pi N}\left|\sum_{k=1}^{N}x_k e^{-ik\omega}\right|^2,$$

称为**周期图**(periodogram).

$I_N(\omega)$可作为谱密度$f_x(\omega)$的估计,且是渐近无偏估计但不是相容估计.

2. 加窗谱估计

对无限长时间的信号作有限时间截取,称加窗处理.加窗谱估计形式为

$$\hat{f}_N(\omega)=\frac{1}{2\pi}\sum_{k=1-N}^{N-1}\hat{r}_k\lambda(k)\mathrm{e}^{-\mathrm{i}k\omega}, \tag{23.53}$$

其中$\lambda(k)$称为时域上的**落后窗**(时滞窗,lag window)或用

$$\hat{f}_x(\omega)=\int_{-\pi}^{\pi}W_N(\omega-\theta)I_N(\theta)\mathrm{d}\theta, \tag{23.54}$$

其中$W_N(\theta)$称为频域上的**谱窗函数**(spectral window function).

二者关系如下:

$$W_N(\theta)=\frac{1}{2\pi}\sum_{k=1-N}^{N-1}\lambda(k)\mathrm{e}^{-\mathrm{i}k\theta},\quad -\pi\leqslant\theta\leqslant\pi, \tag{23.55}$$

$$\lambda_{(k)}=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}k\theta}W_N(\theta)\mathrm{d}\theta,\quad k=0,\pm 1,\pm 2,\cdots,\pm(N-1).$$

谱窗因子的选择原则和$\hat{f}_N(\omega)$具有的重要性质可参看有关文献.

几种常用谱窗函数如下

(1) 截尾窗

$$\lambda(k)=\begin{cases}1, & |k|\leqslant M,\\ 0, & |k|>M,\end{cases}\qquad M<N-1. \tag{23.56}$$

$$W(\theta)=\frac{1}{2\pi}\sum_{k=-M}^{M}\cos k\theta=\frac{1}{2\pi}\frac{\sin\left(M+\frac{1}{2}\right)\theta}{\sin\frac{\theta}{2}}=D_M(\theta).$$

(2) 巴特莱特(Bartlett)窗

$$\lambda(k)=\begin{cases}1-\frac{|k|}{M}, & |k|\leqslant M,\\ 0, & |k|>M.\end{cases} \tag{23.57}$$

$$W(\theta)=\frac{1}{2\pi M}\left(\frac{\sin\frac{M\theta}{2}}{\sin\frac{\theta}{2}}\right)^2=F_M(\theta).$$

(3) 丹尼尔(Daniell)窗

$$\lambda(k)=\frac{\sin\frac{k\pi}{M}}{k\pi/M},$$

$$W(\theta)=\begin{cases}\frac{M}{2\pi}, & -\frac{\pi}{M}\leqslant\theta\leqslant\frac{\pi}{M},\\ 0, & \text{其它}.\end{cases}$$

(4) 一般图凯(Tukey)窗

$$\lambda(k)=\begin{cases}1-2a+2a\cos\frac{\pi k}{M}, & |k|\leqslant M,\\ 0, & |k|>M.\end{cases}\tag{23.58}$$

$$W(\theta)=aD_M\left(\theta-\frac{\pi}{M}\right)+(1-2a)D_M(\theta)+aD_M\left(a+\frac{\pi}{M}\right).$$

(5) 哈明(Hamming)窗

$$\lambda(k)=\begin{cases}0.54+0.46\cos\frac{k\pi}{M}, & |k|\leqslant M,\\ 0, & |k|>M.\end{cases}\tag{23.59}$$

(6) 哈宁(Hanning)窗

$$\lambda(k)=\begin{cases}\frac{1}{2}\left(1+\cos\frac{k\pi}{M}\right), & |k|\leqslant M,\\ 0, & |k|>M.\end{cases}\tag{23.60}$$

(7) 帕森(Parzen)窗

$$\lambda(k)=\begin{cases}1-6\left(\frac{k}{M}\right)^2+6\left(\frac{|k|}{M}\right)^3, & |k|\leqslant\frac{M}{2},\\ 2\left(1-\frac{|k|}{M}\right)^3, & \frac{M}{2}<|k|\leqslant M,\\ 0, & |k|>M.\end{cases}\tag{23.61}$$

(8) 巴特莱特-普利斯特利(Bartlett-Priestley)窗

$$W(\theta)=\begin{cases}\dfrac{3M}{4\pi}\left(1-\left(\dfrac{M\omega}{\pi}\right)^2\right), & |\theta|\leqslant\dfrac{\pi}{M},\\ 0, & |\theta|>\pi/M.\end{cases}\tag{23.62}$$

$$\lambda(k)=\frac{3M^2}{(k\pi)^2}\left(\frac{\sin\dfrac{k\pi}{M}}{\dfrac{k\pi}{M}}-\cos\frac{k\pi}{M}\right).$$

23.10 非平稳时间序列的线性模型

23.10.1 ARIMA 模型

定义 23.10.1 设$\{Z_t\}$ $t=1,2,\cdots$,为一非平稳列,存在正整数 d,使得$\nabla^d Z_t=x_t(t>d)$,是一个 ARMA(p,q)序列,且假定 z_1,$z_2,\cdots,z_d$ 的 $EZ_t=0,EZ_t^2=\sigma^2$ 又与 $t>d$ 后的 x_t 无关,则称$\{Z_t\}$为 d 阶求和序列,d 为求和阶数,相应模型为自回归集成滑动平均模型(autoregressive integrated moving average model)记作 ARIMA(p,d,q),其中∇^d 为 d 阶差分算子:

$$\nabla^d=(1-B)^d=1-C_d^1B+C_d^2B^2+\cdots$$
$$+(-1)^{d-1}C_d^{d-1}B^{d-1}+(-1)^dB^d.$$

这里对 x_t 而言,有 $\varphi(B)x_t=\theta(B)\varepsilon_t$ $(t>d)$.

对 z_t 而言,有 $\varphi(B)\nabla^d z_t=\theta(B)\varepsilon_t$ $(t>d)$.

由于这类求和模型可化为平稳序列模型,故对它的建模、预报等与 ARMA 模型没有本质的不同.

例 23.10.2 ARIMA(0,1,1)模型为 $z_t-z_{t-1}=\varepsilon_t-\theta_1\varepsilon_{t-1}$,$t=2,3,\cdots,N$. 令$\nabla z_t=z_t-z_{t-1}=x_t$,故有 $x_t=\varepsilon_t-\theta_1\varepsilon_{t-1}$是 MA(1)模型,化为 $\varepsilon_t=x_t+\theta_1\varepsilon_{t-1}$ 得

$$S(\theta_1) = \sum_{t=2}^{N} \varepsilon_t^2, \quad \varepsilon_1 = 0.$$

23.10.2 季节性模型

设$\{z_t\}$是一随机序列,若它符合如下形式的模型:

$$\Phi(B^S)\nabla_S^D z_t = \Theta(B^s)A_t, \tag{23.63}$$

则称 z_t 是周期为 s 的季节性序列,此称为**季节性模型**(seasonal model). 其中

$$\nabla_s = 1 - B^s, \quad \nabla_s^D = (1 - B^s)^D,$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_p B^{ps}, \Theta(B^s) = 1 - \Theta_1 B^s - \cdots + \Theta_Q B^{Qs}.$$

若 $D=0$,还应考虑形如

$$\Phi(B^s)Z_t = c + \Theta(B^s)A_t \tag{23.64}$$

的模型. A_t 可以是另一个 ARIMA(p,d,q)序列:

$$\varphi(B)\nabla^d A_t = \theta(B)\varepsilon_t. \tag{23.65}$$

联合式(23.64)与(23.65),可得

$$\varphi(B)\Phi(B^s)x_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{23.66}$$

其中 $x_t = \nabla^d \nabla_s^D z_t$.

式(23.66)称为**乘积型季节性模型**(multiplicative seasonal model),其阶数表示为$(p,d,q)\times(P,D,Q)$.

例 23.10.3 季节性 AR(1)模型为 $z_t - \varphi_{12} z_{t-12} = \varepsilon_t$,其中 ε_t 为白噪声序列,其自相关函数

$$\rho_{12k} = \varphi_{12}^k, \quad k = 1,2,\cdots. \text{其余 } \rho_k = 0.$$

例 23.10.4 乘积型季节性模型$(0,1,1)\times(0,1,1)$为

$$(1 - B^{12})(1 - B)z_t = (1 - Q_1 B)(1 - \theta_2 B^{12})\varepsilon_t.$$

若

$$x_t = (1 - B)(1 - B^{12})z_t,$$

则

$$
\begin{aligned}
x_t &= (1-\theta_1 B)(1-\theta_2 B^{12})\varepsilon_t \\
&= (1-\theta_1 B-\theta_2 B^{12}+\theta_1\theta_2 B^{13})\varepsilon_t \\
&= \varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-12}+\theta_1\theta_2\varepsilon_{t-13}.
\end{aligned}
$$

经计算得

$$
\rho_1=\frac{-\theta_1}{1+\theta_1^2},\ \rho_{12}=\frac{-\theta_2}{1+\theta_2^2},\ \rho_{11}=\rho_{13}=\rho_1\cdot\rho_{12},\text{其余 }\rho_k=0.
$$

24 马尔可夫过程和马尔可夫链

马尔可夫过程(Markov processes)是这样一类随机过程(定义 21.1.6):已知现在的值时,过程的过去与将来独立.此类过程广泛用于描述随时间进化的无后效系统,它最先由俄国数学家 A. A. Марков 研究.通常将参数集和状态空间(定义 21.1.6)都离散的马尔可夫过程称为**马尔可夫链**(Markov chain)或简单地称为**马氏链**.

本章中不特别指明时,所讨论的过程都定义在给定的概率空间上.

24.1 马尔可夫链

24.1.1 定义和例

设 I 为一可列集,$\mathbf{N}=\{0,1,2,\cdots\}$.

定义 24.1.1 取值于 I 的随机变数列 $\{X_n\}_{n\in\mathbf{N}}$ 称为一个**马尔可夫链**(Markov chain).如果对任意 $n\in\mathbf{N}, i_\nu\in I, \nu=0,1,\cdots,n+1$,有

$$P\{X_{n+1}=i_{n+1}|X_n=i_n,\cdots,X_0=i_0\}$$
$$=P\{X_{n+1}=i_{n+1}|X_n=i_n\}.$$

只要上式左方有意义,则称为**马尔可夫性**(Markov property),下式等价于马尔可夫性:

$$P\{X_{n+k}=i_{n+k}|X_n=i_n,X_{\nu(l)}=i_{\nu(l)},\cdots,X_{\nu(1)}=i_{\nu(1)}\}$$
$$=P\{X_{n+k}=i_{n+k}|X_n=i_n\}, \tag{24.1}$$

其中 $l,n,k\in\mathbf{N}/\{0\}$，$\nu(j)\in\mathbf{N}$ 且 $\nu(1)<\nu(2)<\cdots<\nu(l)<n$，$i_{\nu(j)}\in I$，$j=1,\cdots,l$.

例 24.1.2 设有无穷多个瓮，标有号码 0，1，…，第 n 个瓮中放有比例为 p_n 和 q_n 的白球和黑球，$0<p_n<1$，$p_n+q_n=1$. 自第 0 个瓮开始，从中摸出 i_0 个球，其中 i_1 个为白球，接着从第 1 个瓮中摸出 i_1 个球，若有 i_2 个是白球，则从第 2 个瓮中摸出 i_2 个球，一般第 n 个瓮中摸出的白球数即为第 $n+1$ 个瓮中应摸球数，直至摸出白球数为零止. 以 X_n 表示第 n 个瓮中摸出的白球数，显然有

$$\begin{aligned}
&P\{X_{n+1}=j\,|\,X_n=i,X_{n-1}=i_{n-1},\cdots,X_0=i_0\}\\
&\quad=\begin{cases}P_{n+1}^{j}q_{n+1}^{i-j}, & i\geqslant j\\ 0, & i<j\end{cases}\\
&\quad=p\{X_{n+1}=j\,|\,X_n=i\},
\end{aligned}$$

故 $\{X_n\}_{n\in\mathbf{N}}$ 是一个马尔可夫链.

24.1.2 转移概率

定义 24.1.3 对任意 $i,j\in I$，$m,n\in\mathbf{N}$，记

$$\begin{aligned}
{}_mP_{ij}(n)&=P(X_{m-n}=j\,|\,X_m=i),\\
{}_mP(n)&=({}_mP_{ij}(n))_{i,j\in I.}
\end{aligned}\tag{24.2}$$

称 ${}_mP_{ij}(n)$[${}_mP(n)$]为马尔可夫链的 **n 步转移概率**[**n 步转移概率矩阵**](n-steps transition probability[matrixs]). 当 $n=1$ 时简称为**转移概率(矩阵)**，并记成 ${}_mP_{ij}$ 和 ${}_m\boldsymbol{P}$.

性质 24.1.4 由定义知对转移概率成立以下式子：

(1) ${}_mP_{ij}(n)\geqslant 0$， n，m$\in\mathbf{N}$，$i,j\in I$；

(2) $\sum\limits_{j\in I}{}_mP_{ij}(n)=1$， $n,m\in\mathbf{N},i\in I$；

(3) ${}_mP_{ij}(n+n')=\sum\limits_{k\in I}{}_mP_{ik}(n)\cdot{}_mP_{kj}(n')$，

$$m,n,n'\in\mathbf{N},i,j\in I.$$

上式称为查普曼-科尔莫戈罗夫(Chapman-Kolmogorov)方

程，简称 C-K 方程.

一个马尔可夫链的有限维分布族由 X_0 的分布及转移概率矩阵列 $\{_m\boldsymbol{P}\}_{m\in\mathbf{N}}$ 完全决定，前者称为**初始分布**(initial distribution)，后者是马尔可夫链理论中主要研究对象.

24.1.3 齐次马尔可夫链

1. 齐次马尔可夫链定义和存在性

定义 24.1.5 若马尔可夫链 $\{X_n\}_{n\in\mathbf{N}}$ 的转移函数满足：

$$ {}_nP_{ij} = {}_0P_{ij}, \quad n \in \mathbf{N},\ i,j \in I, $$

则称 $\{X_n\}_{n\in\mathbf{N}}$ 为**齐次马尔可夫链**(homogeneous Markov chain). 此时记 $P_{ij}={}_0P_{ij}$，$\boldsymbol{P}=(P_{ij})_{i,j\in I}$，分别称之为齐次马尔可夫链的转移概率和转移矩阵.

以下不特别说明时所考虑的马尔可夫链均是齐次的.

齐次马尔可夫链的有穷维分布由初始分布和转移概率矩阵 $\boldsymbol{P}$ 完全决定. 反之亦然，因为有存在定理.

定理 24.1.6 存在定理 任给概率分布 $\{q_i\}_{i\in I}$ 及满足 $0\leqslant P_{ij}\leqslant 1$ 和 $\sum\limits_{j\in I}P_{ij}=1$，$i,j\in I$ 的矩阵 $\boldsymbol{P}=(P_{ij})_{i,j\in I}$，则存在概率空间 $(\Omega,\mathscr{F},P)$ 及其上的齐次马尔可夫链 $\{X_n\}_{n\in\mathbf{N}}$ 以 I 为状态空间，$\{q_i\}_{i\in I}$ 为初始分布，$\boldsymbol{P}$ 为转移概率矩阵.

2. 高阶转移矩阵的计算

若马尔可夫链的状态空间 I 为有限集，则转移矩阵 $\boldsymbol{P}$ 是有穷矩阵，此时可以通过计算 $\boldsymbol{P}$ 的特征根来求高阶(即多步)转移矩阵 $\boldsymbol{P}(n)$，这在实际问题中很有用.

设 $\boldsymbol{P}=(P_{ij})_{n\times n}$ 有 n 个互不相等的特征根 $\lambda_1,\cdots,\lambda_n$，则 m 阶转移概率

$$ P_{ij}(m) = \sum_{k=1}^{n} A_{ji}(\lambda_k)\lambda_k^m \Big/ \Big(\prod_{\substack{l=1\\ l\neq m}} (\lambda_k - \lambda_l) \Big), \qquad (24.3) $$

$i,j\in I,\quad m=1,2,\cdots,$

其中 $A_{ji}(\lambda_k)$是$(\lambda_k\boldsymbol{E}-\boldsymbol{P})$是第$(j,i)$个元素的代数余子式.

对有重特征根情形亦可推导类似的公式.

3. 状态分类

设马尔可夫链$\{X_n\}_{n\in N}$的转移矩阵为$\boldsymbol{P}=(P_{ij})_{i,j\in I}$，$n$ 阶转移矩阵为$\boldsymbol{P}(n)=\boldsymbol{P}^n=(P_{ij}(n))_{i,j\in I}$.

注记号 24.1.7

$f_{ij}^{(n)}=P(X_{m+n}=j,X_{m+\nu}\neq j,\nu=1,\cdots,n-1|X_m=i).$

$f_{ij}^{(0)}=0,n\geqslant 1,\quad i,j\in I.$

$f_{ij}=\sum_{n=0}^{\infty}f_{ij}^{(n)},u_i=\sum_{n=1}^{\infty}nf_{ii}^{(n)},\quad i,j\in I.$

易见$f_{ij}^{(n)}$为过程自i出发经n步首达状态j的概率，而f_{ij}为经有限步达状态j的概率.

定义 24.1.8 (1) 称状态i为**常返态**(recurent state)，如果$f_{ii}=1$；常返态i称为**正常返**(positive recurent)，若$\mu_i<\infty$；否则称i **零常返**(null recurent).

(2) 若集$\{n:n\geqslant 1\},P_{ii}(n)>0\}\neq\varphi$，称该集合的最大公约数$d=d(i)$为$i$的**周期**(period)，$d>1$时称$i$为**周期的**(periodic).

(3) 称非周期(即$d=1$)正常返态为**遍历状态**(ergodic state).

定理 24.1.9 i 常返$\Leftrightarrow\sum_{n=0}^{\infty}P_{ii}(n)=\infty$；若$i$非常返$\Rightarrow$

$$\sum_{n=0}^{\infty}P_{ii}(n)=\frac{1}{1-f_{ii}}.$$

下面定理可以判别常返态是否为零常返或遍历.

定理 24.1.10

(1) i 常返且有周期d，则$\lim\limits_{n\to\infty}P_{ii}(nd)=d/\mu_i$，

(2) 设i常返，则i零常返$\Leftrightarrow\lim\limits_{n\to\infty}P_{ii}(n)=0$，$i$遍历$\Leftrightarrow\lim\limits_{n\to\infty}P_{ii}(n)=\frac{1}{\mu_i}>0.$

4. 状态空间的分解

利用转移概率的性质可以将马尔可夫链的状态空间分解成一些基本集合,同一集合中状态将具有某些共同的概率特性.

设 $i,j\in I$,若存在 $n\geqslant 1$ 使 $P_{ij}(n)>0$,则记 $i\to j$;如 $i\to j$ 同时 $j\to i$,记为 $i\leftrightarrow j$.

定义 24.1.11 (1) 设 $C\subset I$,若$\forall\ i\in C,j\in I/C,i\to j$ 不成立,则称 C 为**闭集**(closed set). 如单点集$\{i\}$为闭集,称 i 为**吸引态**(absorbing state).

(2) 称状态 i 和 j **同型**(same type),如果它们具有相同的周期并且具有相同的常返性(即同为常返或非常返,常返时同为正常返或同为零常返).

(3) 闭集 C 称为**不可分**(irreducible),若$\forall\ i,j\in C,i\leftrightarrow j$;具有不可分状态空间的马尔可夫链称为**不可分马尔可夫链**(irreducible markov chain).

现将状态空间进行分解.

定理 24.1.12 状态空间可以分解为有限或可列多个互不相交的子集 $D,C_1,C_2,\cdots$之并,使得:

(1) C_n 是常返态组成的不可分闭集,$n\geqslant 1$.

(2) C_n 中的状态同型,$n\geqslant 1$.

(3) D 中状态非常返,且$\forall\ i\in D,j\in C_n,n=1,2,\cdots,j\nrightarrow i$(即 $j\to i$ 不成立). 上述分解是唯一的.

状态空间的这种分解把马尔可夫链分解成一些更基本的链,例如它在 C_n 上的运动就可视为一个以 C_n 为状态空间的马尔可夫链. 从这个观点,不可分马尔可夫链是最基本的链. 下面定理刻画了这种链的运动情况.

定理 24.1.13 周期为 d 的不可分马尔可夫链,其状态空间可以唯一地分解为 d 个互不相交的子集 C_r 之并:

$$C = \bigcup_{r=0}^{d-1} C_r,$$

使自 C_r 中任一状态出发，经一步即转移到 C_{r+1} 中(约定 $C_d = C_0$).

5. 渐近性质

实际应用时常常关心转移次数较大时状态之间的转移情况，因此需考虑 $P_{ij}(n)$ 的极限性态. 我们有

定理 24.1.14 (1) 若 j 非常返或零常返，则 $\forall\, i \in I$

$$\lim_{n\to\infty} P_{ij}(n) = 0.$$

(2) 若 j 正常返，周期为 d，则 $\forall\, i \in I$ 和 $0 \leqslant r \leqslant d-1$，有

$$\lim_{n\to\infty} P_{ij}(nd + r) = f_{ij}(r)\frac{d}{\mu},$$

其中

$$f_{ij}(r) = \sum_{m=0}^{\infty} f_{ij}^{(md+r)}.$$

(3) $\forall\, i, j \in I$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} P_{ij}(k) = \begin{cases} f_{ij}, & j \text{ 正常返}, \\ 0, & j \text{ 非常返或零常返}. \end{cases}$$

6. 平稳分布

定义 24.1.15 设 $(P_{ij})_{i,j\in I}$ 为马尔可夫链 $\{X_n\}_{n\in\mathrm{N}}$ 的转移概率，$\{\pi_k\}_{k\in I}$ 为非负数列，若满足：

$$\sum_{k\in I} \pi_k = 1,$$

$$\sum_{k\in I} \pi_i P_{ik} = \pi_k, \quad k \in I,$$

则称 $\{\pi_k\}_{k\in I}$ 为 $\{X_n\}_{n\in\mathrm{N}}$ 的**平稳分布**(stationary distribution).

通常一个马尔可夫链不一定具有平稳分布，但对一些特殊情形，结论是肯定的，我们有

定理 24.1.16 (1) 有限马尔可夫链的平稳分布恒存在.

(2) 设 $\{X_n\}_{n\in\mathrm{N}}$ 不可分且所有状态都是遍历的，则它有唯一平

稳分布，此分布为

$$\pi_k = \frac{1}{\mu_k} = \lim_{n\to\infty} P_{ik}(n), \quad k \in I.$$

它是下述方程组的唯一解：

$$y_k = \sum_{j\in I} y_j P_{jk}, \quad y_k \geqslant 0, \quad k \in I, \sum_{j\in I} y_j = 1.$$

此定理告诉我们，不可分遍历链的极限分布是平稳的，因此当 n 充分大时可用此平稳分布近似链的一维分布，前者可通过求解方程组得到. 此外 μ_k 也是一个很重要的量，它是自 k 出发再次回到 k 的平均时间，此处再次得到它的计算方法.

24.1.4 在遗传学中的一个应用

马尔可夫链理论广泛应用于物理、化学、工程、生物、医学及经济等领域，这里举一个生物遗传中应用的例.

遗传的一个要素是染色体，每一个生殖细胞只有一组单一的染色体，称为单倍体. 一个后裔分别继承了来自父母的两组染色体，称为二倍体. 遗传性质的携带者称为基因，它们位于染色体上，基因是成对出现的. 一般地成对的基因中每个可以取两种不同形式(等位基因) A 和 a. 在一个总体中基因 A 和 a 的比例是基因频率，记为 p 和 q. 两种等位基因可形成对应于特定位点的三个基因型，AA，Aa 和 aa，AA 个体只产生 A 配子，aa 个体只产生 a 配子，Aa 个体产生数目相等的 A 配子和 a 配子.

考虑一个群体，其中雄性和雌性的基因型频率分布同为：$AA : Aa : aa = d : 2h : r$，$d + 2h + r = 1$. A 和 a 的基因频率为：

$$pt = h + d \text{ 和 } q = h + r.$$

假设配偶是随机形成的且相互独立(称为随机交配)，那么一个后裔具有基因 A 的概率为 p，具有基因型 AA 的概率为 p^2，类似可计算出它具有基因型 Aa 和 aa 的概率分别是 $2pq$ 和 q^2.

为了用马尔可夫链来描述一个给定位点上的遗传过程，用 1，

2,3 表示三种基因型 AA,Aa,aa,用 p_{ij}表示给定一个上代(父与母)的基因型 i 时,后裔出现基因型 j 的概率. 以一对母子为例,设

$$p_{ij}=P(\text{孩子有基因型 } j \mid \text{母亲有基因型 } i)$$

$$i,j=1,2,3.$$

一步转移概率矩阵为

$$\boldsymbol{P}=\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}.$$

可以通过计算相应于频率 d,$2h$,r 的母亲 AA,Aa,aa 的所有可能基因型来确定 p_{ij}. 比如为使孩子有基因型 AA,它必须从其母处以概率 1/2 继承一个 A 基因,并(通过其父亲)从男性群体中以概率 p 得到另一个 A 基因,因此 $p_{21}=\frac{1}{2}P$,类似地得出

$$\boldsymbol{P}=\begin{pmatrix} p & q & 0 \\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}q \\ 0 & p & q \end{pmatrix}$$

由此可以求出二步转移概率矩阵 $\boldsymbol{P}(2)$,它反映从祖父母到孙子、孙女的转移,一般地 $P(n)$反映祖先到其 n 代子孙的转移规律,显然 $P(n)=P^n$,因此:

$$\boldsymbol{P}(n)=\begin{pmatrix} p & q & 0 \\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}q \\ 0 & p & q \end{pmatrix}^n.$$

利用式(24.1)的方法(或用归纳法)得到

$$\boldsymbol{P}(n)=\begin{pmatrix} p^2+pq/2^{n-1} & 2pq+q(q-p)/2^{n-1} & q^2-q^2/2^{n-1} \\ p^2+p(p-q)/2^n & 2pq+(p-q)/2^n & q^2+q(p-q)/2^n \\ p^2-p^2/2^{n-1} & 2pq+p(p-q)/2^{n-1} & q^2+pq/2^{n-1} \end{pmatrix}.$$

显然对 $i=1,2,3$ 都有 $\sum_{n=1}^{\infty} p_{ij}(n)=+\infty$,因此所有状态都是

常返的. 又当 n 充分大后 $p_{ii}(n)p_{ii}(n+1)>0, i=1,2,3$. 故所有状态都是非周期的,由性质 24.1.4(3)及 $p_{ii}(n)\nrightarrow 0(n\to\infty)$,对 $i=1,2,3$ 成立,因此所有状态正常返,从而这是一个不可分的遍历马尔可夫链,直接取极限求得它的唯一平稳分布:

$$\pi_1 = p^2, \quad \pi_2 = 2pq, \quad \pi_3 = q^2.$$

于是得出每个状态的平均返回时间:

$$\mu_1 = \frac{1}{p^2}, \quad \mu_2 = \frac{1}{2pq}, \quad \mu_3 = \frac{1}{q^2}.$$

这说明:一个群体中 A 基因越多,基因型 AA 所需的返回时间的平均值(换代数目)越小,而当 $p=q=1/2$ 时,Aa 型基因的最小期望返回时间 $\mu_{22}=2$(代).

最后由平稳分布定义,对$\forall\, n$,有

$$\pi_i = \pi_1 p_{1i}(n) + \pi_2 p_{2i}(n) + \pi_3 p_{3i}(n), \quad i = 1,2,3,$$

即链的一维分布恒为$(p^2,2pq,q^2)$,换句话说,虽然随机交配一代复一代地进行着,但平稳的基因型分布$(p^2,2pq,q^2)$却永远保持. 这再次肯定了哈代-魏因贝格(Hardy-Weinberg)的平稳定律:不论父母基因型频率$(d,2h,r)$是什么数值,在随机交配的假定下,第一代继承者将有基因型频率$(p^2,2pq,q^2)$,且此频率将永远保持稳定.

24.2 可列状态马尔可夫过程

24.2.1 定义

定义 24.2.1 $(\Omega,\mathscr{F},P)$为一概率空间,定义于其上,以 $T=[0,\infty)$为参数集,以可列集 I 为状态空间的随机过程$\{X_t\}_{t\in T}$称为**马尔可夫过程**(Markov processes),如果它满足:$\forall\, n>1$ 及 $t_i\in T$,$i=1,\cdots,n$, $t_1<t_2<\cdots<t_n$, $i_\nu\in I$, $\nu=1,\cdots,n$,只要 $P(X_{t_\nu}=i_\nu,\nu=1,\cdots,n-1)>0$,就有

$$P(X_{t_n}=i_n|X_{t_\nu}=i_\nu,\nu=1,\cdots,n-1)$$
$$=P(X_{t_n}=i_n|X_{t_{n-1}}=i_{n-1}) \tag{24.4}$$

通常也称状态空间可列的马尔可夫过程为具有**连续参数的马尔可夫链**(Markov chain with continuous parameters).(24.4)式称为**马尔可夫性**.

24.2.2 转移函数

马尔可夫性使得可以用形如$P(X_t=j|X_s=i)$的条件概率来表示过程的有限维分布,这使问题大为简化.类似马尔可夫链,对马尔可夫过程$\{X_t\}_{t\in T}$,引入下面的概念.

以下写$P(X_t=j|X_s=i)$时,表示此条件概率是有定义的.

定义 24.2.2 记 $P_{ij}(s,t)=P(X_t=j\mid X_s=i)$, $i,j\in I,s\leqslant t\in T$,称为马尔可夫过程$\{X_t\}_{t\in T}$的**转移函数**(transition function).而称 $\boldsymbol{P}(s,t)=(P_{ij}(s,t))_{i,j\in I}$ 为**转移函数矩阵**(transition function matrix).

性质 24.2.3

(1) $0\leqslant P_{ij}(s,t)\leqslant 1,\quad i,j\in I,\ s\leqslant t,\ s,t\in T.$

(2) $\sum\limits_{j\in I}P_{ij}(s,t)=1,\quad i\in I,\ s\leqslant t.$

(3) $\forall\ 0\leqslant s\leqslant u\leqslant t,\ s,u,t\in T$ 和 $i,j\in I$,有

$$P_{ij}(s,t)=\sum_{k\in I}P_{ik}(s,u)P_{kj}(u,t). \tag{24.5}$$

(4) $P_{ij}(s,s)=\delta_{ij},\ i,j\in I,\ s\in T.$

(24.5)式也称为查普曼-科尔莫戈罗夫方程,它是马尔可夫过程中的基本方程.

利用转移函数,$\{X_t\}_{t\in T}$的有限维分布表为

$$P(x_0=i_0,x_{t_1}=i_1,\cdots,x_{t_n}=i_n)$$
$$=P(x_0=i_0)\prod_{k=0}^{n-1}P_{i_k},_{i_{k+1}}(t_k,t_{k+1}) \tag{24.6}$$

因此若 x_0 的分布给定，转移函数完全决定马尔可夫过程的有限维分布.

24.2.3 存在性

转移函数(矩阵)的概念也可以独立于马尔可夫过程给出.

定义 24.2.4 称 $I^2\times\{(s,t): s,t\in T, s\leqslant t\}$ 上的实函数 $P_{ij}(s,t)$ 为一个转移函数，若它满足性质 24.2.3 的(1)—(4). 若条件(2)中等号改为不等号"$\leqslant$"，则称 $P_{ij}(s,t)$ 为**次转移函数**(sub-transition function).

显然，一个马尔可夫过程的转移函数是转移函数. 若马尔可夫过程可能在某个时刻从状态空间中消失(也称中断)，则它的转移函数是次转移函数. 对于反面的问题，有下面的定理

定理 24.2.5 设已给 I 上的一个分布 $\{q_i\}_{i\in I}$，即满足 $q_i\geqslant 0$，$\sum\limits_{i\in I}q_i=1$ 的实数列，以及转移函数 $P_{ij}(s,t)(i,j\in I, s,t\in T, s\leqslant t)$，则存在概率空间 $(\Omega,\mathscr{F},P)$ 及定义于其上的马尔可夫过程 $\{X_t\}_{t\in T}$，使 $\{X_t\}_{t\in T}$，以 $\{q_i\}_{i\in I}$ 为初始分布，以 $P_{ij}(s,t)$ 为转移函数.

由于次转移函数可以在状态空间引进附加状态使其成为转移函数，故定理对次转移函数也成立，不过若仅考虑状态空间 I，相应的马尔可夫过程是中断的.

24.2.4 密度函数

虽然转移函数(加上初始分布)完全决定一个马尔可夫过程，但在连续参数情形，不论在理论上还是在实际应用中，要获得一个转移函数是比较困难的(除特殊情形外)，因此转而考虑能(在一定条件下)决定转移函数的有关量. 为此引进下面的概念. 设 $\boldsymbol{P}(s,t)=(P_{ij}(s,t))$ 为转移矩阵.

定义 24.2.6 设对 $i,j\in I$, $\delta\in T$，极限

$$\lim_{\substack{h+k\to 0^+ \\ h=0\text{或}k=0}} \frac{P_{ij}(s-h,s+k)-\delta_{ij}}{h+k} = q_{ij}(s) \tag{24.7}$$

都存在，则称函数矩阵 $\boldsymbol{Q}(s)=(q_{ij}(s))_{i,j\in I}$ 为 $\boldsymbol{P}(s,t)$ 的**密度函数矩阵**(density function matrics).

性质 24.2.7 转移函数矩阵的密度函数矩阵的性质

$$\forall\, i,j\in I,\quad s\in T,$$

$$q_{ij}(s)\begin{cases}\leqslant 0, & \text{如 } i=j,\\ \geqslant 0, & i\neq j.\end{cases}$$

$$\sum_{j\in I} q_{ij}\leqslant 0,\quad i\in I.$$

密度函数具有明确的概率意义. 下面说明这一点.

注记号 24.2.8 $q_i(s)=-q_{ii}(s)$,

$$\sigma_{ij}(s)=\begin{cases}\delta_{ij}, & \text{若 } q_i(s)=0,\\ \dfrac{q_{ij}(s)}{q_i(s)}(1-\delta_{ij}), & q_i(s)\neq 0,\end{cases}$$

则有
$$P_{ii}(s,t)=1-(q_i(s)+o(1))(t-s). \tag{24.8}$$

即 $q_i(s)(t-s)+o(t-s)$ 是 s 时位于状态 i 的条件下时刻 t 不在 i 的概率. 而对 $i\neq j$,

$$\sigma_{ij}(s)=\lim_{t\downarrow s^+}\frac{P_{ij}(s,t)}{1-P_{ii}(s,t)},\quad \text{若 } q_i(s)\neq 0,$$

$\sigma_{ij}(s)$ 可看作下述事件的概率：系统于 s 时位于状态 i，已知发生跳跃，于 t 时刻跳到 j 的概率(当 t 趋近于 s 时).

24.2.5 向前向后方程

由于一个转移函数(矩阵)的密度函数(矩阵)易从实际观测求得，故使人们感兴趣的问题是：任给一个满足性质 24.2.7 的函数矩阵 $\boldsymbol{Q}(s)$(称之为密度函数矩阵)，是否存在转移矩阵 $\boldsymbol{P}(s,t)$ 以 $\boldsymbol{Q}(s)$ 为密度矩阵?在解答这个问题以前自然想到：如何从一转移函数的密度函数反解出该转移函数? 下面的定理在一定程度上给予

回答.

定理 24.2.9 (1) 若对 $i,j\in I,t\in T,P_{ij}(s,t)$作为 s 的函数在 T 上关于 i 一致右连续,(24.7)式中极限关于 j 一致,则

$$\begin{cases}\dfrac{\partial P_{ij}(s,t)}{\partial s}=q_i(s)\left[P_{ij}(s,t)-\sum_{k\in I}\sigma_{ik}(s)P_{kj}(s,t)\right],\\ P_{ij}(s,s)=\delta_{ij},\quad s<t,\ i,j\in I.\end{cases}\tag{24.9}$$

(2) 若 $P_{ij}(s,t)$作为 t 的函数关于 $j\in I$ 在 T 上一致左连续且(24.7)式中极限关于 i 是一致的,且$\sup\limits_{i\in I}q_i(s)<\infty,s\in T$,则有

$$\begin{cases}\dfrac{\partial P_{ij}(s,t)}{\partial t}=-q_j(t)P_{ij}(s,t)+\sum_{k\in I}q_k(t)\sigma_{kj}(t)P_{ik}(s,t),\\ P_{ij}(t,t)=\delta_{ij},\quad s<t,\ i,j\in I.\end{cases}\tag{24.10}$$

上述 q_i 和 σ_{ij}由 24.2.8 定义.

式(24.9)称为**向后方程组**(system of backward equations).式(24.10)称为**向前方程组**(system of forward equations).它们是科尔莫戈罗夫于 1931 年获得的,故常称为科尔莫戈罗夫方程组.

24.2.6 最小过程

设 $\boldsymbol{Q}(s)=(q_{ij}(s))_{i,j\in I}$是满足性质 24.2.7 的函数矩阵,即$\boldsymbol{Q}(s)$是一个密度矩阵,下面的定理说明只需加上自然的限制它就可成为某转移函数的密度.

定理 24.2.10 假设对固定的 $i,j\in I,q_{ij}(s)$是 s 的博雷尔可测函数,且 $q_{ii}(s)$在 T 的任一有限区间上可积,则存在次转移函数矩阵 $\boldsymbol{P}_{\min}(s,t)=(P_{ij}(s,t))_{i,j\in I}$,它以 $\boldsymbol{Q}(s)$为密度矩阵,且满足以下方程:

$$P_{ij}(s,t)=\delta_{ij}+\int_s^t \mathrm{d}u\sum_{k\in I}P_{kj}(u,t)q_{ik}(u),$$

$$P_{ij}(s,t)=\delta_{ij}+\int_s^t \mathrm{d}u\sum_{k\in I}q_{kj}(u)P_{ik}(s,u).$$

此外每一个满足上面方程的次转移函数 $\boldsymbol{P}(s,t)$ 均有 $\boldsymbol{P}(s,t)\geqslant \boldsymbol{P}_{\min}(s,t)$.

$\boldsymbol{P}_{\min}(s,t)$ 称为相应于 $\boldsymbol{Q}$ 的最小(次)转移函数,相应的马尔可夫过程称为**最小过程**(minimal process).

$\boldsymbol{P}_{\min}(s,t)$ 可由 $\boldsymbol{Q}(s)$ 构造,置

$$\varphi_i(s,t)=\exp\left(-\int_s^t q_i(u)\mathrm{d}u\right),$$

$$\hat{q}_{ij}(s)=q_i(s)\sigma_{ij}(s),$$

$$P_{ij}^{(0)}(s,t)=\delta_{ij}\varphi_i(s,t),$$

$$P_{ij}^{(n+1)}(s,t)=\int_s^t \mathrm{d}u\sum_{k\in I}P_{kj}^{(n)}(u,t)\varphi_i(s,u)\hat{q}_{ik}(u),$$

记 $$\boldsymbol{P}^{(n)}(s,t)=(P_{ij}^{(n)}(s,t))_{i,j\in I},$$

$$\boldsymbol{P}^{n}(s,t)=\sum_{k=0}^{n}\boldsymbol{P}^{(k)}(s,t),\quad n=0,1,\cdots,$$

则 $$\boldsymbol{P}_{\min}(s,t)=\lim_{n\to\infty}\boldsymbol{P}^{n}(s,t). \tag{24.11}$$

在实际应用中,大多数情形对给定的 $\boldsymbol{Q}(s)$,相应的转移函数是唯一的,因此常常利用 $\boldsymbol{Q}$ 写出向后或向前方程组,然后求解它得出 $\boldsymbol{P}(s,t)$.

24.2.7 齐次马尔可夫过程

1. 定义和一般性质

定义 24.2.11 具有可列状态空间 I 的马尔可夫过程 $\{X_t\}_{t\in T}$ 称为**齐次马尔可夫过程**(homogenous Markov process),若其转移函数矩阵 $\boldsymbol{P}(s,t)$ 仅依赖于 $t-s$.

于是若记 $\boldsymbol{P}(t)=\boldsymbol{P}(s,s+t)=(\boldsymbol{P}_{ij}(s,s+t))_{i,j\in I}$,

由(24.2.3)得

性质 24.2.12 齐次马尔可夫过程的转移函数的性质：

$$P_{ij}(t) \geqslant 0,\quad i,j \in I,\ t \in T.$$

$$\sum_{j\in I} P_{ij}(t) = 1(\text{中断情形为“}\leqslant\text{”}),\quad i \in I,\ t \in T.$$

$$P_{ij}(s+t) = \sum_{k\in I} P_{ik}(s)P_{kj}(t),\quad i,j \in I,\ s,t \in T.$$

$$P_{ij}(0)=\delta_{ij}.$$

以下考虑的齐次马尔可夫过程，均假定其转移函数满足下述标准性：

$$\lim_{t\to 0^+} P_{ij}(t) = P_{ij}(0) = \delta_{ij},\quad i,j \in I.\qquad(24.12)$$

2. 转移函数的性质

设 $P_{ij}(t)$ $(i,j\in I,t\in T)$是满足性质 24.2.12 中诸式的标准转移函数.

定理 24.2.13 对任意 $i,j\in I$，函数 $P_{ij}(t)$在 $T/\{0\}$上恒为 0 或恒为正.

对齐次马尔可夫过程，它的(转移函数的)密度矩阵(常数阵)在标准性假定下总存在.

定理 24.2.14 $\forall\ i,j\in I$，极限

$$\lim_{t\to 0^+} \frac{P_{ij}(t)-\delta_{ij}}{t} = P_{ij}'(0) = q_{ij}$$

存在，只要转移函数 $P_{ij}(t)$是标准的. 此外对 $i\neq j$，q_{ij}是有限的，而 $\forall\ t>0$，

$$\frac{1-P_{ii}(t)}{t} \leqslant -q_{ii} \triangleq q_i.$$

对于 $P_{ij}(t)$在$(0,\infty)$上的可微性，有

定理 24.2.15 对于标准转移函数 $P_{ij}(t)$，$\forall\ i,j\in I$，$P_{ij}(t)$对 t 在$(0,+\infty)$上具有连续导函数，且

$$P_{ij}'(s+t) = \sum_{k\in I} P_{ik}'(s)P_{kj}(t),\quad s>0,\ t\geqslant 0$$

和

$$P_{ij}'(s+t)=\sum_{k\in I}P_{ik}(s)P_{kj}'(t),\quad s\geqslant 0,\ t>0,$$

这里级数是绝对收敛的. 此外$\forall\ i\in I$,

$$\sum_{j\in I}|P_{ij}'(t)|<\infty,\quad \sum_{j\in I}P_{ij}'(t)=0,\quad t>0,$$

若$q_i<\infty$,$P_{ij}'(t)$在$t=0$也是(右)连续的.

3. 向前向后方程

设$P_{ij}(t)$为标准转移函数,$\boldsymbol{Q}=(q_{ij})$为其密度矩阵,总有

$$P_{ij}'(t)\geqslant\sum_{k\in I}q_{ik}P_{kj}'(t),\quad i,j\in I,\ t\geqslant 0$$

和

$$P_{ij}'(t)\geqslant\sum_{k\in I}P_{ik}(t)q_{kj},\quad i,j\in I,\ t\geqslant 0,$$

若成立等号,就得到向后和向前方程组,这一般需加条件.

定理 24.2.16 (1) 向后方程组成立$\Leftrightarrow\forall\ i\in I$,

$$q_i\triangleq -q_{ii}=\sum_{j\in I/\{i\}}q_{ij}<\infty. \tag{24.13}$$

(2) 若$q_i<\infty$,$i\in I$,且极限

$$\lim_{t\to 0^+}\frac{P_{ij}(k)}{k}=q_{ij},$$

关于$i(\neq j)$一致成立,则向前方程组成立.

注意对I为有限集,(24.13)式,从而向后方程成立,显然向前方程也成立.

在实际应用中,状态空间I常常是有限的,此时向后向前方程的解可以通过计算密度矩阵的特征根和特征向量得出. 于是给出解的两种表达式.

记$I=\{1,2,\cdots,s\}$,

$$\boldsymbol{Q}(\lambda)=\lambda\boldsymbol{E}-\boldsymbol{Q}=(\lambda\delta_{ij}-q_{ij})_{s\times s}$$

为$\boldsymbol{Q}$的特征阵,以$Q_{ij}'(\lambda_l)$表示对应于$\lambda=\lambda_l$的特征阵的第(j,i)个余子式.

定理 24.2.17 (1) 若 $\boldsymbol{Q}$ 有不同的特征值 $\lambda_1,\cdots,\lambda_s$，则向前向后方程满足 $P_{ij}(0)=\delta_{ij}$的解为

$$P_{ij}(t)=\sum_{l=1}^{s}\frac{Q'_{ij}(\lambda_l)e^{\lambda_l t}}{\prod\limits_{\substack{m=1\\m\neq l}}^{s}(\lambda_l-\lambda_m)},\quad i,j=1,2,\cdots,s,t>0.$$

(24.14)

(2) 若 $\boldsymbol{Q}$ 有多重特征根 $\lambda_1,\cdots,\lambda_r$，其重数分别为 $s_1,\cdots,s_r$，$s_1+\cdots+s_r=s$，则向前向后方程满足初始条件 $P_{ij}(0)=\delta_{ij}$的解为：

$$P_{ij}(t)=\sum_{l=1}^{r}\frac{e^{\lambda_l t}}{\prod\limits_{\substack{a=1\\a\neq l}}^{r}(\lambda_l-\lambda_a)^{s_a}}$$

$$\times\sum_{m=0}^{s_l-1}\left[\sum\cdots\sum\prod_{\substack{a=1\\a\neq l}}^{r}\binom{s_a+m_a-1}{m_a}(\lambda_a-\lambda_l)^{-m_a}\right]$$

$$\times\left[\sum_{n=0}^{s_l-1-m}\frac{t^{s_l-1-m-n}}{(s_{l-1-m-n})!n!}\frac{d^n}{d\lambda_l^n}Q_{ij}'(\lambda_l)\right],\qquad(24.15)$$

其中 $\sum\cdots\sum$ 表示 $r-1$ 个和式，每一个对应一个 $a\neq l,l=1,\cdots,r$，并加遍 $m_a=0,\cdots,m$，这里 $m=(m_1+\cdots+m_r)-m_l$.

用上述公式时，除计算特征值外，还须求出相应特征阵的代数余子式. 利用下面定理，也可通过特征值和特征向量表达.

定理 24.2.18 (1) 式(24.13)的解为(矩阵形式)：

$$\boldsymbol{P}(t)=\boldsymbol{T}(k)\boldsymbol{E}(t)\boldsymbol{T}^{-1}(k),$$

其中 $\boldsymbol{E}(t)=(e^{\lambda_i t}\delta_{ij})_{s\times s}$，$\boldsymbol{T}$ 是由相应于 $\lambda_1,\cdots,\lambda_s$ 的 s 个特征向量 $\boldsymbol{T}_1,\cdots,\boldsymbol{T}_s$ 为列向量构成的矩阵，满足 $\boldsymbol{T}^{-1}\boldsymbol{QT}=(\lambda_i\delta_{ij})_{s\times s}$.

(2) 式(24.14)给出的解可由下列表达：

$$\boldsymbol{P}(t)=\boldsymbol{T}\,\mathrm{diag}(\boldsymbol{E}_1(t),\cdots,\boldsymbol{E}_r(t)]\boldsymbol{T}^{-1},$$

其中 $\mathrm{diag}(E_1(t),\cdots,E_r(t)]$表示一个 $s\times s$ 矩阵，它的主对角线元

素自左上角开始依次为 s_1 个 $e^{\lambda_1 t}$,…,s_r 个 $e^{\lambda_r t}$,其余元素为 0. $\boldsymbol{T}=[T_1^{(0)},\cdots,T_1^{(s_1-1)};T_2^{(0)},\cdots,T_2^{(s_2-1)};\cdots;T_r^{(0)},\cdots,T_r^{(s_r-1)}]$.

这里 $T_i^{(j)}$ 如下得到:取 $\boldsymbol{Q}(\lambda)=(\lambda\boldsymbol{E}-\boldsymbol{Q})$ 的一个非零特征向量

$$(Q_1(\lambda),Q_2(\lambda),\cdots,Q_s(\lambda))^{\mathrm{T}}$$

置

$$T_i^{(k)}=\frac{1}{k!}\frac{\mathrm{d}^k}{\mathrm{d}\lambda^k}(Q_1(\lambda),\cdots,Q_s(\lambda))^{\mathrm{T}}|_{\lambda=\lambda_i},\quad k=0,\cdots,s_i-1$$

即可.

4. Q 过程

和非齐次情况一样,实际问题中常常关心的是:已给密度矩阵 $\boldsymbol{Q}$,即其元素满足 $0\leqslant q_{ij}<\infty,i\neq j,\ \sum_{j\neq i}q_{ij}\leqslant -q_{ii}\leqslant\infty$ 的矩阵,是否存在转移函数(或次转移函数)$P_{ij}(t)(i,j\in I,t\in T)$使 $\boldsymbol{Q}$ 为 $\boldsymbol{P}(t)=(P_{ij}(t))$的密度?若存在是否唯一?它们满足向后和向前方程组否?这些问题对满足 $-q_{ii}<\infty$ 的 $\boldsymbol{Q}$,得到了较为完满的结果. 首先由定理 24.2.10,对应于 $\boldsymbol{Q}$ 的次转移函数(简称为 **Q 过程**(Qprocess))存在,且可由(24.11)构造出**最小 Q 过程**(minimal Q process),它满足向后和向前方程组. 对唯一性有下面诸定理.

对 $\lambda>0$,以 $\mathscr{M}_\lambda(\mathscr{M}_\lambda^+)$ 和 $\mathscr{L}_\lambda(\mathscr{L}_\lambda^+)$分别表示方程组(向量形式)

$$\lambda\boldsymbol{U}-\boldsymbol{Q}\boldsymbol{U}=0$$

和

$$\lambda\boldsymbol{V}-\boldsymbol{V}\boldsymbol{Q}=0$$

的解(非负解)全体,$\mathscr{M}_\lambda^+$ 和 $\mathscr{L}_\lambda^+$ 的维数记为 m^+ 和 n^+.

定理 24.2.19 下列条件等价

(1) 满足向后方程组的 Q 过程唯一.

(2) 对某 $\lambda>0$,$\mathscr{M}_\lambda$ 仅为零解.

(3) 对某 $\lambda>0$,$\mathscr{M}_\lambda^+$ 仅为零解.

定理 24.2.20 (1) 若最小过程不中断(即 $P_{\min}(t)$为转移函

数),或者它虽中断(为次转移函数)但 $n^{-}=0$,则满足向前方程组的 Q 过程唯一.

(2) 若最小解中断且 $n^{+}=1$,则有无穷多个 Q 过程满足向前方程,其中只有一个是不中断的.

(3) 若最小解中断且 $n^{+}>1$,则有无穷多个 Q 过程满足向前方程,且有无穷多个是不中断的.

一般地有以下结论:

定理 24.2.21 Q 过程的唯一性准则(uniqueness criterion for Q process) 假定 $q_{ii}>-\infty, i\in I$,则 Q 过程唯一的充要条件是最小 Q 过程不中断,或最小过程中断并满足下面两个条件:

(1) $$\inf_{i\in I}\lambda\sum_{j\in I}\Phi_{ij}(\lambda)=\eta_\lambda>0,\ \lambda>0,$$

其中 $$\Phi_{ij}(\lambda)=\int_0^\infty e^{-\lambda t}P_{ij}(t)dt.$$

(2) $n^{+}=0$,即 $\mathscr{L}_\lambda^{+}$ 只含零解.

5. 生灭过程

定义 24.2.22 定义在概率空间 $\langle\Omega,\mathscr{F},P\rangle$ 上的齐次马尔可夫过程称为**生灭过程**(birth and death process),若其状态空间 $I=N$,且其转移函数的密度矩阵 $Q=(q_{ij})_{i,j\in N}$ 具有下述形式:

$$q_{ij}=0,\quad 若\ |i-j|>1;$$
$$q_{i,i+1}=b_i,\quad q_{i,i-1}=a_i,$$
$$q_{ii}=-(a_i+b_i),$$

其中 $a_i>0, b_i>0$ 为实数,$i\geqslant 1, a_0=0$.

生灭过程广泛应用于各个领域,详细的讨论可参考文献[40]和[36]以及第 32 章.

24.3 一般状态空间上的马尔可夫过程

设 $\langle E,\mathscr{B}\rangle$ 为完备可分距离空间,$\mathscr{B}$ 为博雷尔 σ 域,考虑 E 上

的马尔可夫过程.

24.3.1 马尔可夫过程及其转移函数

定义 24.3.1 设$(\Omega,\mathscr{F},P)$为一概率空间,$(\mathscr{F}_t)_{t\geqslant 0}$为$\mathscr{F}$的上升子$\sigma$代数族(即满足$\mathscr{F}_t\subset\mathscr{F}_{t+s}\subset\mathscr{F},s\geqslant 0$的$\sigma$代数族).定义在$(\Omega,\mathscr{F},P)$上的$E$值随机过程$\{X_t\}_{t\geqslant 0}$称为马尔可夫过程,如果对任意$t\geqslant 0$,$X_t$为$\mathscr{F}_t$可测,且

$$P(X_t\in\Gamma|\mathscr{F}_s)=P(X_t\in\Gamma|X_s)\qquad(\text{a.s.p})\quad(24.16)$$

对$0\leqslant s\leqslant t$和$\Gamma\in\mathscr{B}$成立.

为进一步研究马尔可夫过程,引进下面

定义 24.3.2 实值函数$P(s,x;t,\Gamma)$,$0\leqslant s\leqslant t$,$x\in E$,$\Gamma\in\mathscr{B}$,称为一个转移函数,如

(1) $\forall\ 0\leqslant s\leqslant t$,$P(s,x,t,0)$为$\mathscr{B}$上的概率测度;

(2) $\forall\ 0\leqslant s\leqslant t$,$\Gamma\in\mathscr{B}$,$P(s,0;t,\Gamma)$为$E$上的$\mathscr{B}$可测函数;

(3) $\forall\ 0\leqslant s\leqslant t\leqslant u$,$\Gamma\in\mathscr{B}$,

$$P(s,x;u,\Gamma)=\int P(t,y;u,\Gamma)p(s,x;t,\mathrm{d}y)\qquad(24.17)$$

(24.17)式称为查普曼-科尔莫戈罗夫方程或C-K方程.

若$P(s,x;t,\Gamma)=P(t-s,x,\Gamma)$仅与$t-s$有关,$0\leqslant s\leqslant t$,则称为齐次转移函数.

设$\{X_t\}_{t\geqslant 0}$为一马尔可夫过程,如存在转移函数$P(s,x;t,\Gamma)$使

$$P(X_t\in\Gamma|X_s)=P(s,X_s;t,\Gamma)\qquad(\text{a.s.p})$$

就称$P(s,x;t,\Gamma)$为$\{X_t\}_{t\geqslant 0}$的转移函数,转移函数为齐次的马尔可夫过程称为齐次马尔可夫过程.

一个马尔可夫过程的转移函数和初始分布(X_0的分布)完全决定它的有限维分布族.

24.3.2 存在性定理

定理 24.3.3 设$P(s,x;t,\Gamma)$为一转移函数,μ为$\mathscr{B}$上任一

概率测度,则存在一概率空间$(\Omega,\mathscr{F},P)$及其上的E值随机过程$\{X_t\}_{t\geq 0}$,它以μ为初始分布,关于$\mathscr{F}_t^X=\sigma(X_s,s\leqslant t)$为马尔可夫过程,$P(s,x;t,\Gamma)$为转移函数.

以下将不区分马尔可夫过程与它的转移函数,且主要考虑齐次马尔可夫过程.

24.3.3 停时和强马尔可夫过程

设$(\Omega,\mathscr{F},P)$为一概率空间,$(\mathscr{F}_t)_{t\geq 0}$为$\mathscr{F}$的一族上升子$\sigma$代数.

定义 24.3.4 称映射$\tau:\Omega\to[0,+\infty)$为$(\mathscr{F}_t)$**停时**(stopping time)或**马尔可夫时间**,若对任意$t\geqslant 0$,$\{\omega:\tau(\omega)\leqslant t\}\in\mathscr{F}_t$.

设τ为$(\mathscr{F}_t)$停时,定义$\mathscr{F}_\tau\doteq\{A:A\in\mathscr{F},A(\tau\leqslant t)\in\mathscr{F}_t,t\geqslant 0\}$,则$\mathscr{F}_\tau$为一$\sigma$代数.

定义 24.3.5 设$\{X_t\}_{t\geq 0}$为$(\Omega,\mathscr{F},P)$上的$\{\mathscr{F}_t\}$齐次马尔可夫过程,$P(t,x,\Gamma)$为其转移函数,若$\{X_t\}_{t\geq 0}$循序可测,且对每一a.s.有限停时τ,$\forall\ t\geqslant 0$和$\Gamma\in\mathscr{B}$有

$$P(X_{\tau+t}\in\Gamma\,|\,\mathscr{F}_\tau)=P(t,X_\tau,\Gamma)\qquad(\text{a.s.p})$$

则称$\{X_t\}_{t\geq 0}$为$(\mathscr{F}_t)$**强马尔可夫过程**(strong Markov process).

24.3.4 半群及其无穷小算子

考虑齐次马尔可夫过程.

虽然转移函数$P(t,x,\Gamma)$(加上初始分布)完全决定一个马尔可夫过程,但实际中要获得一个具体的转移函数是很困难的,因此用转移函数来描述马尔可夫过程是不方便的,必须寻求其它有关的量.马尔可夫过程理论中这方面的一个途径就是用巴拿赫(Banach)空间上算子半群及其无穷小生成算子来描述,导致这种方法的是以下事实:

设$B(E)$为E上有界$\mathscr{B}$可测函数全体,以上确界作范数后

$B(E)$成为巴拿赫空间. 设 $P(t,x,\Gamma)$ 为转移函数, $\forall t\geqslant 0$, $f\in B(E)$, 令

$$T_t f(x)=\int f(y)P(t,x,\mathrm{d}y),$$

则$\{T_t\}_{t\geqslant 0}$为 $B(E)$上一族**压缩算子半群**(semigroup of contraction operators), 即 $\|T_t\|\leqslant 1,\forall t\geqslant 0$ 且 $T_tT_s=T_{s+t},\forall s,t\geqslant 0$, 显然此算子半群完全决定了 $P(t,x,\Gamma)$. 通常称上述$\{T_t\}_{t\geqslant 0}$为**转移半群**(transition semigroup).

$B(E)$上压缩半群$\{T_t\}$的无穷小算子 A 由下式决定:

$$Af=\lim_{t\to 0}\frac{T_tf-f}{t}, \qquad (24.18)$$

其中收敛为按范数收敛. A 为 $B(E)$上线性算子, 其定义域是使(24.18)中极限存在的一切 $f\in B(E)$, 记之为 D_A. 在一定条件下, 压缩半群为其无穷小算子唯一决定, 因此可用无穷小算子描述马尔可夫过程.

在实际问题中, 决定无穷小算子的有关参数常可由观测求得, 例如在很多场合, 过程的无穷小算子是一个二阶微分算子, 此时它的系数可由某些物理量决定. 无穷小算子不但刻画出过程的性质, 有时还可由它直接构造出过程的轨道, 这方面的知识可参见第 25 和 30 章.

24.3.5 跳跃马尔可夫过程和费勒过程

跳跃马尔可夫过程(Markov jump processes)和**费勒过程**是最典型而重要的两类过程. 我们来讨论它们对应的半群和无穷小算子.

1. 跳跃马尔可夫过程

设 $\lambda(x)$为 $B(E)$中的非负函数, $\mu(x,\Gamma)$是定义在 $E\times\mathscr{B}$ 上的实函数, 对固定的 Γ, $\mu(\cdot,\Gamma)$为 $\mathscr{B}$ 可测, 固定 $x\in E$, $\mu(x,\cdot)$是

$\mathscr{B}$ 上概率测度. 记

$$Af(x) = \lambda(x)\int(f(y) - f(x))\mu(x,\mathrm{d}y), \quad f \in B(E),$$

则 A 为 $B(E)$上的有界线性算子,它是跳跃过程典型的无穷小算子,事实上可以构造一跳跃过程以 A 为无穷小算子.

2. 费勒过程

为叙述简单设 E 为紧距离空间,以 $C(E)$表示 $B(E)$中连续函数全体,视为 $B(E)$的子空间.

定义 24.3.6 称 $C(E)$上的正强连续压缩半群$\{T_t\}_{t\geq 0}$为**费勒半群**(Feller semi-group),若 $T_t=0$.

若$\{T_t\}$是由转移函数 $P(t,x,\Gamma)$产生的,若 T_t 保持 $C(E)$不变,则它是费勒半群.

定理 24.3.7 设 E 为紧距离空间,$\{T_t\}_{t\geq 0}$为 $C(E)$上的费勒半群,μ 为 $\mathscr{B}$ 上任一概率测度,则存在概率空间$(\Omega,\mathscr{F},P)$及其上的 E 值马尔可夫过程$\{X_t\}_{t\geq 0}$,以 μ 为初始分布且对$\forall\ f\in C(E)$,有 $E(f(x_{s+t})\,|\,\mathscr{F}_s^X=T_tf(x_s)$,此外 X 是具有右连续轨道的强马尔可夫过程.

关于费勒半群的无穷小算子,有下面一般结果.

定理 24.3.8 E 为紧距离空间,A 为 $C(E)$上线性算子,则 A 的闭包 $\bar{A}$ 为单值且产生一个正强连续压缩半群$\Leftrightarrow$

(1) D_A 稠于 $C(E)$;

(2) A 满足正极大原理;

(3) 存在 $\lambda>0$,使 $\lambda-A$ 的值域稠于 $C(E)$.

称费勒半群对应的马尔可夫过程为费勒过程(Feller process).

24.3.6 马尔可夫过程与鞅

任给一马尔可夫过程,下面定理使它联系一个鞅(定义

29.1.1).

定理 24.3.9 设$\{X_t\}_{t\geqslant 0}$为$(\Omega,\mathscr{F},P)$上的马尔可夫过程，其无穷小算子是A，则对任意$f\in D_A$，过程$\left\{f(X_t)-\int_0^t Af(X_s)\mathrm{d}s\right\}_{t\geqslant 0}$在$P$下关于$\{\mathscr{F}_t^X\}$为鞅.

下面说明，一定条件下的马尔可夫过程可以用鞅予以描述.先引进一些术语.

定义 24.3.10 (1) 设$A\subset B(E)\times B(E)$，A的鞅问题解是指：定义在某个概率空间$(\Omega,\mathscr{F},P)$上的E值随机过程$\{X_t\}_{t\geqslant 0}$在P下关于$({}^*\mathscr{F}_t^X)$是一个鞅，这里${}^*\mathscr{F}_t^X=\mathscr{F}_t^X\vee\sigma\left\{\int_0^s h(X_u)\mathrm{d}u,s\leqslant t,h\in B(E)\right\},t\geqslant 0$.

(2) μ为$\mathscr{B}$上的概率测度，(A,μ)鞅问题解是指以μ为初始分布的A鞅问题解.

(3) (A,μ)鞅问题的唯一性是指：若$\{X_t\}$和$\{Y_t\}$是(A,μ)鞅问题的两个解，则它们的有限维分布相同.

以$D_E[0,\infty)$表示$[0,\infty)$右连左极E值函数全体，φ_E为其上由柱集产生的σ代数，$\{X_t\}_{t\geqslant 0}$为$D_E[0,\infty)$上坐标过程，于是一个A鞅问题解就指$(D_E[0,\infty),\varphi_E)$上的一个概率测度$P$.此外以$\mathscr{T}$表示$[0,\infty)$上非负有界可测函数全体，$\mathscr{F}_t=\mathscr{F}_t^X\vee\sigma\{(\tau\leqslant s):s\leqslant t,t\in\mathscr{T}\}$，$\Gamma$为$\varphi_E$上概率测度的一个集合，对$\mathscr{B}$上的概率测度$\nu$，记$\Gamma_\nu=\{P\in\Gamma:PX_0^{-1}=\nu\}$.

定理 24.3.11 设E为完备可分紧距离空间，假设对$\mathscr{B}$上任一概率测度ν，$\Gamma_\nu\neq\varnothing$，记$A_0=\left\{(f,g)\in B(E)\times B(E):\forall\ p\in\Gamma,f(X_t)-\int_0^t g(X_s)\text{ 为}(\mathscr{F}_t)\text{ 鞅}\right\}$，则

(1) 存在一个线性耗散(dissipative)算子$A\supset A_0$，使$\lambda-A$的

值域为 $B(E)(\lambda>0)$，D_A 的有界弱闭包在 $B(E)$ 中稠.

(2) 对 $\mathscr{B}$ 上任一概率测度 ν，存在 $P_\nu\in\Gamma_\nu$，它是 (A,ν) 鞅问题的唯一解.

(3) A 的每一鞅问题满足

$P(X_{\tau+t}\in C|\mathscr{F}_t)=P_{X_\tau}(C)$，$C\in\varphi_E$，$\tau\in\mathscr{T}$，$P_X$ 为 (A,δ_X) 的鞅问题解，$x\to P_x$ 是 $\mathscr{B}$ 可测的.

此外还有

定理 24.3.12 设 E 可分，$A\subset B(E)\times B(E)$ 是线性的且为耗散；又设存在 $A'\subset A$，A' 线性且 $\lambda-A'$ 的值域的闭包为 $\overline{D}_A=L$（对某 $\lambda>0$），若 L 具有性质：对 $\mathscr{B}$ 上的测度 μ,ν，$\forall f\in L$，$\mu f=\nu f\Rightarrow\mu=\nu$，那么对 $\mathscr{B}$ 上任意概率测度 μ，(A,μ) 鞅问题的解是马尔可夫过程，且鞅问题唯一性成立.

这里没有区分 $B(E)$ 上的算子与其图象，A 还可以为多值的.

用鞅刻画马尔可夫过程时，对算子 A 的限制较无穷小算子弱得多，因此鞅方法是一种有力的工具. 有关这方面的细节和严格的陈述参见文献[54].

25 扩散过程

25.1 扩散过程的重要性与引例

25.1.1 扩散过程的重要性

本章将研究系统状态以连续方式变化的一类重要的马尔可夫过程.历史上这类过程的研究起源于对扩散现象的研究.

许多物理、化学、生物、经济和社会现象可以用扩散过程(diffusion process)作模型来研究.例如相互作用条件下的可列多个质点的分子运动、理想市场中的价格波动、某些有噪声干扰的通讯系统、有刺激的神经活动、群体增长中的变异、在竞争和其它相互关系下的物种数量的变化、进化中的基因取代等等.

扩散过程与其它随机过程关系密切,因而其理论研究也相当重要.作时间尺度变换和把状态变量重新规格化,许多马尔可夫过程可用扩散过程来近似.

一维扩散过程的许多重要泛函(像首次通过时间等)的概率分布、边界性质及平稳分布等,其计算常归纳为解有简单边界条件的二阶微分方程,因而能精确求出.这些公式推广到多维扩散过程也常无困难,只是求解时要遇到偏微分方程,因此实际找出其精确解,常非易事.

25.1.2 引例

下例说明一个在数轴整数点 Z_1 上作随机游动(random walk)的质点是如何经极限过渡为在全直线 R^1 上的扩散的.这种方法称为**扩散近似**(approximation by diffusion).

设一质点在某时刻位于 i，则下一单位时刻以概率 $p(\geqslant 0)$ 向右移到 $i+1$，而以概率 $q \triangleq 1-p(\geqslant 0)$ 向左移到 $i-1, i\in Z_1$. 若以 $X(n)$ 表示 n 个单位时刻后（也称 n 步后）质点的位移，则易知 $\{X(n), n=0,1,2,\cdots\}$ 是齐次马尔可夫链，其状态空间 $E=Z_1$. 一步转移概率为

$$p_{i\ i+1}=p,\quad p_{i\ i-1}=q,\quad p_{ij}=0,\quad j\neq i\pm 1,\quad i\in E.$$

令

$$X_i=\begin{cases}1, & \text{质点第 } i \text{ 步向右移,}\\ -1 & \text{质点第 } i \text{ 步向左移,}\end{cases}$$

则质点每一步的平均位移 $EX_i=p-q$，方差为 $DX_i=4pq$. 由假设知 $X(n)=\sum_{i=1}^{n}X_i$，故 $EX(n)=n(p-q), DX(n)=4npq$.

今考虑极限过程，假定：(1)每步的空间长度为 Δx；(2)每步所需时间长度为 Δt. 于是 $(0,t]$ 时间内共有 $t/\Delta t$ 步转移，从而

$$\text{平移位移} \doteq \frac{t}{\Delta t}(p-q)\Delta x=t(p-q)\frac{\Delta x}{\Delta t},$$

$$\text{位移方差} \doteq 4\frac{t}{\Delta t}pq(\Delta x)^2=4tpq\frac{(\Delta x)^2}{\Delta t}.$$

欲 $\Delta t\to 0$ 时上述极限方差存在，应要求 $\lim\limits_{\Delta t\to 0}(\Delta x)^2/\Delta t$ 存在. 此时如又要求平均位移有限，则应存在 $\lim\limits_{\Delta t\to 0}(p-q)/\Delta x$. 为此令

$$\frac{(\Delta x)^2}{\Delta t}=2D,\quad p=\frac{1}{2}+\frac{C}{2D}\Delta x,\quad q=\frac{1}{2}-\frac{C}{2D}\Delta x, \tag{25.1}$$

其中 C 和 $D(>0)$ 为常数. 从而

$$\begin{cases}(0,t] \text{ 中平均位移} \triangleq m(t)=2Ct,\\ (0,t] \text{ 中位移方差} \triangleq \sigma^2(t)=2Dt.\end{cases} \tag{25.2}$$

$(0,t]$ 中总位移可视为独立同分布的小位移之和，其中每一位移取 $\pm\Delta x$ 之一. 令 $\Delta t\to 0$，由中心极限定理及(25.2)知，总位移的

极限分布密度函数为

$$f(t,x) = \frac{1}{\sqrt{2\pi}\sigma(t)}\exp\{-(x-m(t))^2/2\sigma^2(t)\}$$
$$= \frac{1}{2\sqrt{\pi Dt}}\exp\{-(x-2Ct)^2/4Dt\}. \qquad (25.3)$$

它描述了在给定假设下，质点在全直线上“扩散”的概率规律.

25.2 扩散过程的定义

25.2.1 利用转移概率定义

以 $X(t)$表时刻 t 质点的位置，它是随机变量. 设 $X(t)=x$，考察在时间间隔$(t,t+\Delta t]$中状态变化 $\Delta X(t) \triangleq X(t+\Delta t)-X(t)$. 设过程连续(21.3.3)，故对任 $\delta>0$ 有

$$\lim_{\Delta t\to 0} P(|\Delta X(t)| > \delta | X(t) = x) = 0,\text{或}$$
$$\lim_{\Delta t\to 0} \int_{|y-x|>\delta} \mathrm{d}_y F(t,x;t+\Delta t,y) = 0, \qquad (25.4)$$

这里 $F(t,x;t+\Delta t,y) \triangleq P(X(t+\Delta t)\leqslant y | X(t)=x)$，且设上一极限对 x 为一致的，本节以下的极限也恒设如此. 由于物理和生物现象中的扩散，其无穷小位移的均值与方差多与 Δt 成正比，故可设

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t}\int_{-\infty}^{\infty} (y-x)\mathrm{d}_y F(t,x;t+\Delta t,y)$$
$$= \lim_{\Delta t\to 0} \frac{1}{\Delta t} E(\Delta X(t) | X(t) = x) \triangleq a(t,x), \qquad (25.5)$$
$$\lim_{\Delta t\to 0} \frac{1}{\Delta t}\int_{-\infty}^{\infty} (y-x)^2\mathrm{d}_y F(t,x;t+\Delta t,y)$$
$$= \lim_{\Delta t\to 0} \frac{1}{\Delta t} E[(\Delta X(t))^2 | X(t) = x] \triangleq b(t,x) > 0. \qquad (25.6)$$

连续性条件(25.4)常强化为邓肯条件(Denkin condition)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x|>\delta} \mathrm{d}_y F(t,x;t+\Delta t,y) = 0, \qquad (25.7)$$

或更强的

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x|>\delta} (y-x)^2 \mathrm{d}_y F(t,x;t+\Delta t,y) = 0. \quad (25.8)$$

显见，当(25.8)成立时，(25.6)与(25.7)可分别写为截尾矩的形式：

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x|>\delta} (y-x) \mathrm{d}_y F(t,x;t+\Delta t,y) = a(t,x), \qquad (25.9)$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|y-x|>\delta} (y-x)^2 \mathrm{d}_y F(t,x;t+\Delta x,y) = b(t,x) > 0. \qquad (25.10)$$

应注意，(25.9)与(25.10)存在不保证 F 有一阶矩和二阶矩.

定义 25.2.1 称满足(25.7)，(25.9)和(25.10)的马尔可夫过程为**扩散过程**(diffusion process)，式中 $a(t,x)$ 称为**漂移参数**(drift parameter)，**无穷小均值**(infinitesimal mean)或**无穷小期望位移**(expected infinitesimal displacement). 称 $b(t,x)$ 为**扩散参数**(diffusion parameter)，**无穷小方差**(infinitesimal variance).

注 25.2.2 利用转移概率，扩散过程尚有如下定义(其中限制(1)最强，理论讨论中更常用(2))：

(1) 满足(25.5)，(25.6)和(25.10)的马尔可夫过程，称为扩散过程.

(2) 连续参数的马尔可夫过程，如其几乎所有样本轨道 $X(t)$ 是 t 的连续函数，则称其为扩散过程 70，且常设过程规则 70.

相空间 E 可易为某有限或无限区间. 在某边界点上如 $b(t,x)=0$，或 $a(t,x)$，$b(t,x)$ 不存在，则称此扩散过程为**奇异的**(singular)，又，不失一般性几乎所有轨道连续的过程总可设为一

切样本轨道为连续的(见[24]).

对于齐次扩散过程,(25.5)和(25.6)可分别写为

$$\lim_{t\downarrow 0}E_x(X(t)-x)/t$$
$$=\lim_{t\downarrow 0}E(X(t)-x|X(0)=x)/t=a(x), \tag{25.11}$$
$$\lim_{t\downarrow 0}E_x((X(t)-x)^2)/t$$
$$=\lim_{t\downarrow 0}E((X(t)-x)^2|X(0)=x)/t=b(x)>0. \tag{25.12}$$

回忆(25.1)中例子.由无限制随机游动过渡得到的极限过程,易验证它是 $a(t,x)=2C,b(t,x)=2D$ 的齐次扩散过程.

每一个连续马尔可夫过程,都可以看作为一个不连续马尔可夫过程的极限情形.

25.2.2 利用无穷小算子定义

注记号 25.2.3 D_2^x:在 x 邻域内二次连续可微实函数类. $D_{\mathscr{U}}^x=\{f:\in C(E),\mathscr{U}f(x)\triangleq\lim_{n\to+\infty}[E_xf(X_{h_n})-f(x)]/E_xh_n$ 存在$\}$,其中 h_n 为 x 的邻域 U_n 的首达时,$U_n\downarrow\{x\}$.

定义 25.2.4 设 $\mathscr{U}$ 为$(R,\mathscr{B})$上连续强马尔可夫过程(定义 24.3.5)X 的特征算子(定义 21.6.10),如对任何 $x\in E,D_2^x\subset D_{\mathscr{U}}^x$,则称 X 为**扩散过程**.

性质 25.2.5 扩散过程的存在性 设函数 $a(x)$和 $b(x)$满足

(1) 有界且赫尔窦(Hölder)条件成立:存在 $k,\alpha>0$ 使 $|f(x)-f(y)|\leqslant k|x-y|^\alpha,x,y\in R,f=a$ 或 b;

(2) 存在常数 $\rho>0$ 使 $b(x)\geqslant\rho,x\in E$.

则在$(R,\mathscr{B})$上存在扩散过程使其特征算子 $\mathscr{U}$ 满足 $\mathscr{U}f(x)=a(x)f'(x)+\frac{1}{2}b(x)f''(x),f\in D_2^x$.

性质 25.2.6 两种定义间的关系 设 X 是$(R^1,\mathscr{B})$中齐次马尔可夫连续过程,且对每 $x\in R^1$

(1) 存在邻域 U_x 使 $E_x e_{U_x}<\infty$,其中 e_U 是首次流出 U 的时间:$e_U=\inf(t:X_t\notin U)$;

(2) 存在极限

$$\lim_{t\downarrow 0}\frac{1}{t}\int_R (y-x)P(t,x,\mathrm{d}y)=a(x),$$

$$\lim_{t\downarrow 0}\frac{1}{t}\int_R (y-x)^2P(t,x,\mathrm{d}y)=b(x),$$

$a(x)$,$b(x)$连续,且极限号下的比关于 $x\in R,t\geqslant 0$ 一致有界,则 X 是(齐次)的扩散过程,$a(x)$为漂移系数而 $b(x)$为扩散系数.

以下除另有说明,凡扩散过程均由(25.4)定义.

25.3 科尔莫戈罗夫向前向后方程

25.3.1 科尔莫戈罗夫向前向后方程

定理 25.3.1 科尔莫戈罗夫向前方程(Kolmogorov forward equation) 设 $X=(X_t)_{t\in\Delta}$,$\Delta=[a,b]$是扩散过程(25.2.1),其转移概率密度函数 $p(s,x;t,y)$存在,且下列偏导数存在、连续:$\frac{\partial}{\partial t}p(s,x;t,y)$,$\frac{\partial}{\partial y}[a(t,y)p(s,x;t,y)]$,$\frac{\partial^2}{\partial y^2}[b(t,y)p(s,x;t,y)]$.则 $p(s,x;t,y)$满足下列偏微分方程.

$$\frac{\partial}{\partial t}p(s,x;t,y)=-\frac{\partial}{\partial y}[a(t,y)p(s,x;t,y)]+\frac{1}{2}\frac{\partial^2}{\partial y^2}[b(t,y)p(s,x;t,y)] \quad (25.13)$$

定理 25.3.2 科尔莫戈罗夫向后方程(Kolmogorov backward equation) 设 $X=(X_t)_{t\in\Delta}$,$\Delta=[a,b]$是扩散过程(25.2.1),其转移概率分布函数 $F(s,x;t,y)\triangleq P(X_t\leqslant y\mid X_s=x)$ 的下列偏导数存在且连续:$\frac{\partial}{\partial x}F(s,x;t,y)$与$\frac{\partial^2}{\partial x^2}F(s,x;t,y)$,则

$F(s,x;t,y)$满足方程

$$\frac{\partial}{\partial s}F(s,x;t,y) = -a(s,x)\frac{\partial}{\partial x}F(s,x;t,y) - \frac{1}{2}b(s,x)\frac{\partial^2}{\partial x^2}F(s,x;t,y) \quad (25.14)$$

当 $p(s,x;t,y)$存在时，可以 p 易(25.14)中之 F.

在特殊场合，福克-普朗克(Fokker-Planck)首先得到过它们，因此也常称它们为福克-普朗克方程.

齐次时(25.13)与(25.14)分别有如下形式

$$\frac{\partial}{\partial t}p(t,x,y) = -\frac{\partial}{\partial y}[a(y)p(t,x,y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(y)p(t,x,y)], \quad (25.15)$$

$$\frac{\partial}{\partial t}F(t,x,y) = a(x)\frac{\partial}{\partial x}F(t,x,y) + \frac{1}{2}b(x)\frac{\partial^2}{\partial x^2}F(t,x,y), \quad (25.16)$$

初始条件是 $F(0^+,x,y)=\delta(y-x)$，即$=1$，如 $x\leqslant y$；$=0$，反之. 可以 $p(t,x,y)$易(25.16)中之 $F(t,x,y)$，如能移密度存在且满足相应的条件.

25.3.2 科尔莫戈罗夫方程解的存在与唯一性

关于科尔莫戈罗夫扩散方程解的存在性与唯一性，没有初等的处理方法. 在一定条件下费勒给出了理论证明，但大部分近代结果都是以半群理论为依据的.

注 对扩散过程$(X_t)_{t\in T}$，$T=(l,r)\subset R_1$，费勒(1952年)把求解与边界分类(25.5)联系起来得到如下结果(下述"正则"、"自然"及"流出"等概念见式(25.5).

(1) 当两边界没有一个是正则时，尽管科尔莫戈罗夫扩散方程的初值问题可能有许多解，但两方程只有一个公共的基本解(或格林函数).

(2) 两边界都是自然边界时，两方程初值问题都有唯一的解，由公共的基本解给出.

(3) 两边界至少有一正则时，两方程有无穷多个公共的基本解.

(4) 当 l 自然、r 流出时，向后方程的初值问题有无穷多解，而向前方程的初值问题是唯一确定的.

(5) 当 l 自然而 r 正则时，两方程的初值问题都有无穷多解.

25.3.3 科尔莫戈罗夫方程的解法

科尔莫戈罗夫方程是抛物型偏微分方程，这类方程最有用的方法是分离变量法与拉普拉斯变换法(或相当的特征函数法).

(1) 分离变量法：由于 X 在方程中任意固定，故齐次转移密度函数 $p(t,x,y)$ 可简记 $f(t,y)$. 如 $f(t,y)$ 可分离变量，即可写 $f(t,y)=g(y)T(t)$，则代入科尔莫戈罗夫方程即可得到关于 t 和关于 y 的两个常微分方程，从而求解.

(2) L-变换法：对 $p(s,x;t,y)$ 作拉普拉斯变换：

$$\text{对 } t, \quad p^*(s,x;\lambda,y) = \int_0^{\infty} p(s,x;t,y)\mathrm{e}^{-\lambda t}\mathrm{d}t,$$

$$\text{对 } s, \quad p_*(t,x;\lambda,y) = \int_0^{\infty} p(\lambda,x;t,y)\mathrm{e}^{-\lambda s}\mathrm{d}s.$$

对向前[后]科尔莫戈罗夫方程，两边对 $t[s]$ 作拉普拉斯变换，得到关于 $y[x]$ 的非齐次二阶常微分方程. 解这个常微分方程，并将所得解反演求出 $p(s,x;t,y)$. 在一些简单情况可以查拉普拉斯变换表(参阅附表 10)，否则要利用复变函数论中的围道积分技术.

25.3.4 几类扩散过程的科尔莫戈罗夫方程

由下一定理,立即可判断本段所列的几类过程是一扩散过程.

定理 25.3.3 扩散基本变换定理(elementary transformation theorem for diffusion) 设$\{X(t),t\geqslant 0\}$是注25.2.2(2)中的扩散过程,且规则相空间为区间I,其端点为l和$r(>l)$.又设$\{X(t)\}$有无穷小参数$\mu(x)$和$\sigma^2(x)$.令g为I上严格单调函数且在I内有连续二阶导数,则$Y(t)=g[X(t)]$定义一个值在以$g(l)$和$g(r)$为端点的区间上的规则扩散过程,且有无穷小参数

$$\begin{cases}\mu_Y(y)=\dfrac{1}{2}\sigma^2(x)g''(x)+\mu(x)g'(x),\\ \sigma_Y^2(y)=\sigma^2(x)[g''(x)]^2.\end{cases}\tag{25.17}$$

其中$y=g(x)$.

(1) 标准布朗运动(见定义26.1.2)$(B_t)_{t>0}$的科尔莫戈罗夫向后方程是

$$\frac{\partial p}{\partial t}=\frac{1}{2}\frac{\partial^2 p}{\partial x^2},\quad t>0,\ -\infty<x<\infty.\tag{25.18}$$

在$t=0$的适当初值条件下$\left(\int_0^y p(t,x,z)\mathrm{d}z\to\delta(y-x),t\downarrow 0\right)$有唯一转移密度解,它是高斯核(kernel).

$$\varphi(t,x,y)=\frac{1}{\sqrt{2\pi t}}\exp\{-(y-x)^2/2t\},$$
$$t>0,\ x,y\in(-\infty,\infty)\tag{25.19}$$

(2) **有漂移的布朗运动**(Brownian motion with drift) $W_t=\sigma B_t+\mu t$,它是无穷小参数$a(x)=\mu,b(x)=\sigma^2$的扩散过程,其向后方程是

$$\frac{\partial p}{\partial t}=\mu\frac{\partial p}{\partial x}+\frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial x^2},\quad t>0,\ -\infty<x,y<\infty.\tag{25.20}$$

在 $t=0$ 的适当初值条件下有唯一的转移概率密度解

$$p(t,x,y)=\varphi(\sigma^2 t, x+\mu t, y). \qquad (25.21)$$

(3) 奥恩斯坦-乌伦贝克(Ornstein-Uhlenbech)**过程**$(V_t)_{t>0}$,是 $a(x)=-\beta x$, $b(x)=\alpha$ 的扩散过程,其向后方程是

$$\frac{\partial p}{\partial t}=-\beta x\frac{\partial p}{\partial x}+\frac{1}{2}\alpha\frac{\partial^2 p}{\partial x^2},$$

$$t>0,\ -\infty<x,\ y<\infty. \qquad (25.22)$$

唯一的概率密度解是$\left(\text{在边界条件} \lim\limits_{y\to\pm\infty} yp=0=\lim\limits_{y\to\pm\infty}\frac{\partial}{\partial y}p \text{ 之下}\right)$.

$$p(t,x,y)=\varphi\left(\frac{\alpha}{2\beta}(1-e^{-2\beta t}), xe^{-\beta t}, y\right). \qquad (25.23)$$

(V_t)与布朗运动(B_t)间关系是

$$V(t)=e^{-\beta t}B(\sigma^2(e^{2\beta t}-1)/2\beta).$$

(4) 节 25.1 中无限制游动的扩散近似,其向前方程为

$$\frac{\partial f}{\partial t}=-2C\frac{\partial f}{\partial y}+p\frac{\partial^2 f}{\partial y^2},\quad f(t,y)=p(t,0,y). \qquad (25.24)$$

其唯一的概率密度解

$$f(t,y)=\frac{1}{2\sqrt{\pi Dt}}\exp\{-(y-2Ct)^2/(4Dt)\} \qquad (25.25)$$

25.4 平稳分布、首中时及其泛函

25.4.1 平稳分布

设有无穷小参数 $a(x)$, $b(x)$, $I=(l,r)$中齐次的扩散过程,有转移密度且平稳分布存在,则平稳密度 $\psi(x)$满足

$$\psi(y)=\int\psi(x)p(t,x,y)\mathrm{d}x,\quad t>0. \qquad (25.26)$$

ψ 满足的微分方程是

$$0=\frac{1}{2}\frac{\partial^2}{\partial y^2}[b(y)\psi(y)]-\frac{\partial}{\partial y}[a(y)\psi(y)].\qquad(25.27)$$

解出的平稳分布是

$$\begin{aligned}\psi(x)&=C_1\frac{S(x)}{s(x)b(x)}+C_2\frac{1}{s(x)b(x)}\\&=m(x)[C_1S(x)+C_2],\end{aligned}\qquad(25.28)$$

其中 $s(x)$,$S(x)$及 $m(x)$见定义 25.5.3 和定义 25.5.4,常数 C_1,C_2 由"$\psi(x)$为概率密度函数"这一条件求出:$\psi(x)\geqslant 0,x\in(l,r)$;$\int_l^r\psi(x)\mathrm{d}x=1$. 如这样的 C_1 与 C_2 找不到,则平稳密度不存在.

25.4.2 首中时及其泛函

令 $X=(X_t)_{t\in I}$,$I=(l,r)$为齐次规则扩散过程,无穷小参数 $a(x)$与 $b(x)>0$ 连续,令

$T(x)=\inf\{t:X_t=x,t\geqslant 0\}$(点 x 的**首中时**(hitting time))

例 25.4.1 首中时的几个有关泛函

(1) $u(x)\triangleq P(T(b)<T(a)\mid X_0=x)$,$l<a<x<b<r$,满足

$$\begin{cases}0=a(x)\dfrac{\mathrm{d}u}{\mathrm{d}x}+\dfrac{1}{2}b(x)\dfrac{\mathrm{d}^2u}{\mathrm{d}x^2},\quad a<x<b,\\ u(a)=0,\quad u(b)=1.\end{cases}$$

解得
$$u(x)=\frac{S(x)-S(a)}{S(b)-S(a)},\quad a\leqslant x\leqslant b.$$

(2) $v(x)\triangleq E[T^*\mid X_0=x]$, $l<a<x<b<r$,其中 $T^*=T(a)\wedge T(b)$. v 满足

$$\begin{cases}-1=a(x)\dfrac{\mathrm{d}v}{\mathrm{d}x}+\dfrac{1}{2}b(x)\dfrac{\mathrm{d}^2v}{\mathrm{d}x^2},\\ v(a)=v(b)=0.\end{cases}$$

解得

$$v(x)=2\Big\{u(x)\int_x^b[S(b)-S(y)]m(y)\mathrm{d}y$$

$$+ [1-u(x)]\int_a^x [S(y)-S(a)]m(y)\mathrm{d}y\Big\}.$$

(3) $w(x) \triangleq E\left[\int_0^{T^*} g(X_s)\mathrm{d}s \mid X_0 = x\right], l<a<x<b<r$,其中 g 是有界连续函数. 当 $g(x)$ 表示该过程在状态 x 所需的费用率,则$A \triangleq \int_0^{T^*} g(X_s)\mathrm{d}s$ 将是直到跑出(a,b)时为止的总费用,取 $g(x)\equiv 1$,则 $A = T^*$.

$w(x)$满足

$$\begin{cases} -g(x) = a(x)\dfrac{\mathrm{d}w}{\mathrm{d}x} + \dfrac{1}{2}b(x)\dfrac{\mathrm{d}^2 w}{\mathrm{d}x^2}, \\ w(a) = w(b) = 0. \end{cases}$$

解得

$$\begin{aligned} w(x) &= 2\Big\{u(x)\int_x^b [S(b)-S(y)]m(y)g(y)\mathrm{d}y \\ &\quad + [1-u(x)]\int_a^x [S(y)-S(a)]m(y)g(y)\mathrm{d}y\Big\} \\ &= \int_a^b G(x,y)g(y)\mathrm{d}y, \end{aligned}$$

其中

$$G(x,y) = \begin{cases} 2\dfrac{[S(x)-S(a)][S(b)-S(y)]}{S(b)-S(a)}\dfrac{1}{b(y)s(y)}, & a\leqslant x\leqslant y\leqslant b, \\ 2\dfrac{[S(b)-S(x)][S(y)-S(a)]}{S(b)-S(a)}\dfrac{1}{b(y)s(y)}, & a\leqslant y\leqslant x\leqslant b. \end{cases}$$

称 $G(x,y)$为过程 X 在$[a,b]$上的格林函数.

上述诸式中之 $S(x)$,$s(x)$及 $m(x)$见定义 25.5.3 和 25.5.4.

25.5 规则扩散过程的边界分类

科尔莫戈罗夫方程不能唯一决定扩散过程,要考虑边界条件.

25.5.1 四个函数

定义 25.5.1 如对任一 $t\geqslant 0$，$P_x(X_t=x)\triangleq P(X_t=x|X_0=x)=1$，则称此点 $x\in E$ 为 X 的**吸收点**(absorbing point)。如 x 为吸收点，则对一切 t 和 $f\in B(E)$ 有 $T_t f(x)=E_x f(X_t)=f(x)$.

如 x 不是吸收点，则存在 x 的邻域 U，使 $E_y e_U<\infty$，一切 $y\in U$. 这里 e_U 为首次流出 U 的时间：$e_U=\inf(t:X_t\notin U, t\geqslant 0)$.

定义 25.5.2 设 X 是区间 I 上的扩散过程，I 的两端点 $l<r$，均可开可闭，也可取到无穷. 设 $x\in(l,r)$，如对任 $y\in(l,r)$，$P_x(\tau_y<\infty)>0$，则称 x 是过程 X 的**规则点**(requalr point)，这里 τ_y 是 X 首中 y 的时刻. 如 X 的每一状态均为规则，则称 X 为**规则过程**(regular process).

定义 25.5.3 尺度函数与尺度测度　X 的无穷小位移和方差分别为 $a(x)$ 和 $b(x)>0$. 令

$$s(y)=\exp\left\{-\int_{y_0}^{y}[2a(x)/b(x)]\mathrm{d}x\right\},$$

$$S(x)=\int_{x_0}^{x}s(y)\mathrm{d}y,\quad x_0,y_0\in(l,r).$$

称 $S(x)$ 为 X 的**尺度函数**(scale function).

对 $J=[c,d]\subset(l,r)$，令 $S[J]=S(d)-S(c)$. 由集函数 $S[J]$ 产生的测度称为**尺度测度**(scale measure).

定义 25.5.4 速度函数与速度测度　由**速度密度**(speed density) $m(x)=1/[b(x)s(x)]$ 决定集函数 $M[J]=\int_c^d m(x)\mathrm{d}x$，由此产生的测度称为**速度测度**(speed measure).

定义 25.5.5 四个函数 $S(l,x]$，$\Sigma(l)$，$M(l,x]$ 和 $N(l)$

$$S(l,x]\triangleq\lim_{c\downarrow l}S[c,x],\quad l<x<r.$$

$$\Sigma(l)\triangleq\int_l^x M[y,x]\mathrm{d}s(y).$$

$$M(l,x]=\lim_{a\downarrow l}M[a,x].$$

$$N(l)\triangleq\int_l^r S[y,x]\mathrm{d}M(y)=\int_l^x M(l,y]\mathrm{d}S(y).$$

粗略地说，$\Sigma(l)$是度量到达边界点 l 所花费的时间，或是从内点 $x<b$ 出发到达内点 b 的时间，$M(l,x]$度量 X 在 l 附近的速度而 $N(l)$是从边界 l 出发到达内点 $x\in(l,r)$所用的时间.

定义 25.5.6 如 $S(l,x]<\infty$，任 $x\in(l,r)$，则称边界 l 为**吸引的**(attracting). 如 $\Sigma(l)<\infty$，则称 l 为**可达的**(attainable)，反之称**不可达的**(unattainable).

对边界 r 可仿上定义 $S[x,r)$，$\Sigma(r)$等等，进行分类.

25.5.2 边界分类

按四个函数是否无穷而对边界 l 分类. 由于四个函数的内在关系($N(l)<\infty\Rightarrow M(l,x]<\infty$，$\Sigma(l)<\infty\Rightarrow S(l,x]<\infty$以及 $\Sigma(l)+N(l)=S(l,x]\times M(l,x]$)，只有六种可能的组合. 在美国通行的费勒分类，与前苏联学者采用的略有不同.

边界 l 分类表如表 25.1.

表 25.1 费勒分类及前苏联分类比较

<table>
<tr><th colspan="4">判别</th><th colspan="4">边界术语</th></tr>
<tr><th>$S(l,x]$</th><th>$M(l,x]$</th><th>$\Sigma(l)$</th><th>$N(l)$</th><th>费勒</th><th colspan="3">前苏联</th></tr>
<tr><td>$<\infty^*$</td><td>$<\infty^*$</td><td>$<\infty$</td><td>$<\infty$</td><td colspan="2">规则</td><td rowspan="3">吸引</td><td rowspan="2">可达</td></tr>
<tr><td>$<\infty$</td><td>$=\infty^*$</td><td>$<\infty^*$</td><td>$=\infty$</td><td colspan="2">流出</td></tr>
<tr><td>$<\infty^*$</td><td>$=\infty^*$</td><td>$=\infty^*$</td><td>$=\infty$</td><td rowspan="3">自然
($\Sigma(l)=\infty^*$
$N(l)=\infty^*$)</td><td>吸引且不可达</td><td rowspan="4">不可达</td></tr>
<tr><td>$=\infty^*$</td><td>$=\infty^*$</td><td>$=\infty$</td><td>$=\infty^*$</td><td rowspan="3">自然
($S(l,x]=\infty^*$</td><td rowspan="3">不吸引</td></tr>
<tr><td>$=\infty^*$</td><td>$=\infty^*$</td><td>$=\infty$</td><td>$=\infty$</td></tr>
<tr><td>$=\infty^*$</td><td>$<\infty$</td><td>$=\infty$</td><td>$<\infty^*$</td><td>流入</td></tr>
</table>

表中 * 号表示该行成立的一个起码的充分条件.

边界大致性态：

(1) **规则边界**(regular boundary)　规则边界既可进入也可离开. 当 $M[\{l\}]=\infty$, l 为吸引; $M[\{l\}]=0$ 时为反射; 而 $0<M[\{l\}]<\infty$ 时边界为一粘性壁, 质点在边界上有一个严格正的持续时间, 只是这持续时间不含任何区间.

另一可能是质点在首达 l 时重新开始这个状态空间在内部的过程. 此过程可以是马尔可夫过程, 却因轨道不处处连续而不是扩散过程. 此过程的转移密度满足向后方程.

(2) **流出边界**(exit boundary)　向着边界的漂移可以是这样的, 使得边界自动地起吸收壁作用, 因此不能加上边界条件去解科尔莫戈罗夫扩散方程. 马尔可夫性有可能保留. 如保留, 在 l 处过程的持续时间必遵从指数分布, 以一定的概率规律, 一步跳进 (l,r) 内.

(3) **流入边界**(entrance boundary)　从状态空间内部不可能到达一个流入边界, 但从此边界出发时却很快跑到区间内部, 并且不再回到该流入边界.

(4) **费勒自然边界**(Feller natural boundary)　这样一个扩散过程既不能在有限平均时间内到达边界, 也不能自一个自然边界开始. 因此状态空间中可略去费勒意义下的自然边界点, 在扩散方程的求解时不必在端点附加边界条件.

25.6 几个例子与应用

作为扩散过程一个重要特例的布朗运动过程(也称维纳-列维(Wiener-Levy)过程), 及其有关的过程(有吸引或反射的布朗运动、有漂移的布朗运动、几何布朗运动及布朗桥过程)的定义与应用, 见节 26.1 和节 26.4.

1. **奥恩斯坦-乌伦贝克过程**可作为**爱伦菲斯特**(Ehrenfest)**模**

型的扩散近似，也可作一个物理过程的模拟.

(1) 爱伦菲斯特模型刻画在$(-N,N)$间整数点上的如下游动：$p_{i\,i+1}=\frac{1}{2}\left(1+\frac{i}{N}\right)$，$p_{i\,i-1}=\frac{1}{2}\left(1-\frac{i}{N}\right)$，而在边界有弹射：$p_{N\,N-1}=1=p_{-N\,-N+1}$. 仿 25.1.2 引例但选 Δt 和 Δx 使$\frac{(\Delta x)^2}{\Delta t}=\alpha$ 及 $N\Delta t=\beta^{-1}$，$\beta>0$. 则极限过渡得到齐次扩散过程，无穷小位移 $a(x)=-\beta x$，无穷小方差 $b(x)=\alpha$.

(2) 一个自由质点在开始$(t=0)$时速度为 x，确定 t 后速度介于 y 与 $y+\mathrm{d}y$ 之间的概率. 奥恩斯坦-乌伦贝克得到的速度的转移密度 $p(t,x,y)$应满足$\frac{\partial p}{\partial t}=\beta\frac{\partial(yp)}{\partial y}+\frac{\alpha}{2}\frac{\partial^2 p}{\partial y^2}$. 其中 $\beta=f/m$，$\alpha=2fkT/m^2$，f 是摩擦系数，m 是质点的质量，k 为粘滞系数，T 绝对温度. 这是 $a(x)=-\beta x$，$b(x)=\alpha$ 的扩散过程的向前方程，其解见(25.23)，其均值函数 $m(t)=x\mathrm{e}^{-\beta t}$，方差 $\sigma^2(t)=\frac{\alpha}{2\beta}(1-\mathrm{e}^{-2\beta t})$. 当 $t\to\infty$时 $p(t,x,y)$趋于正态密度$\frac{1}{\sqrt{\pi\alpha/\beta}}\exp\{-\beta y^2/\alpha\}$. 如 X_0 有此正态分布密度，则 X_t 的分布密度也是它，即上述分布有平稳分布性质.

2. 遗传模型及有关扩散近似

(1) **赖特-费西尔(Wright-Fisher)模型**　设一群体大小为常数 N，其中有 i 个 A 型个体和 $N-i$ 个 a 型个体. 一个 A 型个体一次生育中转换为一个 a 型的概率为 α，而 a 型转换为 A 型概率为 β. 如 A 型和 a 型能活到生育下一代的能力之比为 $1+s$ 和 1，$s>0$，于是选择 A 型较优于 a 型. 于是在再生育之前生出 A 型个体的期望比率为 $p_i=(1+s)p_A/[(1+s)p_A+sp_a]$，这里 $p_A=i(1-\alpha)+(N-i)\beta$，$p_a=i\alpha+(N-i)(1-\beta)$.

赖特-费西尔模型假定次一代的构成是由 N 个二项试验所决定，而每一个二项试验中转换为一个 A 型个体的概率为 p_i，则群

体过程$\{X_t \triangleq$第t代中A型个体数$, t \geqslant 0\}$，是一马尔可夫过程，转移概率$p_{ij} = \binom{N}{j} p_i^j (1-p_i)^{N-j}$.

对甚大的N，该过程可以几个扩散过程来近似：

(i) 设**变种效果**(mutation effects)$\alpha = \frac{\gamma_1}{N} > 0, \beta = \frac{\gamma_2}{N} > 0, s = 0$，则在生育时间$[N\tau]$上群体中$A$型个体比例$Y_N(\tau) = X([N\tau])/N$，这里$Y_N(\tau)$中的$\tau = 1$，对应$X_t$过程的传至$N$代所需的时间，则$\{Y_N(\tau), \tau \geqslant 0\}$之极限过程是状态空间为$[0,1]$的一个扩散过程，其无穷小系数为$a(y) = -\gamma_1 y + \gamma_2(1-y)$和$b(y) = y(1-y)$.

(ii) 设$\alpha = \beta = 0$而$s = \sigma/N$，过程$Y_N(\tau) = X([N\tau])/N$，在N足够大时可视为$a(y) = \sigma y(1-y), b(y) = y(1-y)$的扩散过程.

(iii) 设$0 < \alpha, \beta < 1$而$s = 0$，则对大的N，过程$V_N(t) = (X(t) - N\beta/(\alpha+\beta))(N\alpha\beta/(\alpha+\beta)^2)^{-\frac{1}{2}}$可以无穷小系数$a(x) = -(\alpha+\beta)x$和$b(x) = 1$的一个奥恩斯坦-乌伦贝克过程近似.

(2) **纯生过程**(pure birth process)的扩散近似. 设群体里的各个个体生育相互独立，在时刻t时群体大小为x，在$(t, t+\mathrm{d}t)$时段内生出一新个体的概率为$\lambda(x)\Delta t + o(\Delta t)$，群体大小变为$x+1$以外的概率为$o(\Delta t)$，不发生大小变化的概率为$1 - \lambda(x)\Delta t + o(\Delta t)$. 这是群体增长的纯生过程. 当群体相当大时这一非连续过程可视为连续的. 特别当$\lambda(x) \equiv \lambda > 0$时可能的一个扩散近似是有无穷小参数$a(x) = \alpha x, b(x) = \beta x$的齐次扩散过程，向前方程为

$$\frac{\partial f(t,x)}{\partial t} = -\alpha \frac{\partial [xf(t,x)]}{\partial x} + \beta \frac{\partial^2 [xf(t,x)]}{\partial x^2}, \quad 0 < x < \infty.$$

应用费勒边界分类判据可知$x = 0$为流出而$x = \infty$是自然的. 因此在$x = 0$处无边界条件可加，初值$f(t,0)$则任意.

令$X_0 = x_0, \alpha \neq 0$，则费勒(1939年，1951年)证明解是

$$p_t(x_0,x)=\frac{\alpha}{\beta(e^{\alpha t}-1)}\left[\frac{x_0e^{\alpha t}}{x}\right]^{1/2}J_1\left(\frac{2\alpha[x_0e^{\alpha t}]^{1/2}}{\beta(e^{\alpha t}-1)}\right)$$

$$\times\exp\left\{\frac{-\alpha[x_0e^{\alpha t}+x]}{\beta(e^{\alpha t}-1)}\right\},$$

其中$J_1(\cdot)$为第一类贝塞尔函数. 又$EX_t=x_0e^{\alpha t}$.

(3) 核子级联过程,在均匀核材料中的级联是由一个单位能量的核子(中子或质子)贯穿原子核引起的. 这一贯穿导致更多核子生成,它们转而与介质的其它原子核碰撞. 令$F_n(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n;t)d\varepsilon_1d\varepsilon_2,\cdots,d\varepsilon_n$为均匀核材料的深度$t$处,发现有$n$个位于能量比区间$(\varepsilon_i,\varepsilon_i+d\varepsilon_i)$内的核子,而其它地方没有核子的微分概率,$\varepsilon_i=E_i/E_0$, $i=1,2,\cdots,n$. E_0为原核子的能量,E_i为次级核子能量. Messel证明F_n满足n维扩散方程

$$\frac{\partial F_n}{\partial t}=-nF_n+\sum_{i\neq j}\int_0^1F_{n-1}(\varepsilon_1{}',\varepsilon_2{}',\cdots,\varepsilon_{n-2}{}',\varepsilon;t)$$

$$W\left(\frac{\varepsilon_j}{\varepsilon},\frac{\varepsilon_i}{\varepsilon}\right)\varepsilon^{-2}d\varepsilon.$$

其中$\varepsilon_1{}',\cdots,\varepsilon_{n-2}{}'$是不含$\varepsilon_i$及$\varepsilon_j$的能量比,而$W(\varepsilon',\varepsilon'')d\varepsilon'd\varepsilon''dt$是一个能量为$E_0$的原粒子在$(t,t+dt)$内受一次碰撞产生一些位于能量区间$(E',E'+dE')$与$(E'',E''+dE'')$内的次级核子的概率,这里$\varepsilon'=E'/E_0$, $\varepsilon''=E''/E_0$,而W是齐次函数. 此方程解为

$$F_n=\frac{2^{n-1}}{(n-1)!}L_n\ \Lambda_n(\varepsilon_1,\cdots,\varepsilon_n)f_n(t).$$

其中

$$L_n=\frac{1}{(2\pi i)^n}\int_{c_1-i\infty}^{c_1+i\infty}\cdots\int_{c_n-i\infty}^{c_n+i\infty}\varepsilon_1^{-s_1+1}\cdots\varepsilon_n^{-s_n+1}ds_1\cdots ds_n,$$

$$\Lambda_n=\prod_{k=n-1}^{1}\sum{}'W(s_k,s_{k+1}+\cdots+s_n),$$

$$W(s_i,s_j)=\int_0^1\int_0^{1-\varepsilon_j}\varepsilon_i^{s_i}\varepsilon_j^{s_j}W(\varepsilon_i,\varepsilon_j)d\varepsilon_id\varepsilon_j.$$

求和$\sum'$表示对$s_1,s_2,\cdots,s_k$, $\sum\limits_{j=k+1}^{n}s_j$这$k+1$个变量中任选二个所

得的一切组合求和.

$$f_n(t) = \mathrm{e}^{-t}(1 - \mathrm{e}^{-t})^{n-1}.$$

25.7 多维扩散过程

该 n 维随机向量过程(stochastic vector process) $X(t) \triangleq (X_1(t), X_2(t), \cdots, X_n(t)), t \geqslant 0$ 的状态空间是 n 维欧氏空间 R^n，或它的一个开区域，概率转移分布函数是

$$\begin{aligned}F(t,x;t+\Delta t,y) &\triangleq F(t,x_1,\cdots,x_n;t+\Delta t,y_1,\cdots,y_n)\\&= P(X_i(t+\Delta t) \leqslant y_i,\\&\qquad i = 1,2,\cdots,n \mid X_i(t) = x_i, i = 1,2,\cdots,n).\end{aligned}$$

这里 $x = (x_1, x_2, \cdots, x_n), y = (y_1, y_2, \cdots, y_n) \in R^n$.

以 S_δ 记以 x 为圆心、半径 $\delta > 0$ 的球，连续性条件(25.4)现在是

$$\lim_{\Delta t \to 0} \int_{R^n \setminus S_\delta} \mathrm{d}_y F(t,x;t+\Delta t,y) = 0. \tag{25.29}$$

代替(25.5)和(25.6)，分别是

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{R^n} (y_j - x_j) \mathrm{d}_y F(t,x;t+\Delta t,y)\\&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E(X_j(t+\Delta t) - X_j(t) \mid X(t) = x)\\&\triangleq a_j(t,x), \quad j = 1,2,\cdots,n.\end{aligned} \tag{25.30}$$

$$\begin{aligned}&\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{R^n} (y_i - x_j)(y_j - x_j) \mathrm{d}_y F(t,x;t+\Delta t,y)\\&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[(X_i(t+\Delta t) - X_i(t))(X_j(t+\Delta t)\\&\quad - X_j(t)) \mid X(t) = x]\\&= b_{ij}(t,x), \quad i,j = 1,2,\cdots,n.\end{aligned} \tag{25.31}$$

其中 $B(t,x) = (b_{ij}(t,x))_{n \times n}$ 正定. (25.29)可强化为

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{R^n\setminus S_\delta}\mathrm{d}_y F(t,x;t+\Delta t,y)=0. \qquad (25.32)$$

及更强的

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{R^n\setminus S_\delta}(y_i-x_i)(y_j-x_j)\mathrm{d}_y F(t,x;t+\Delta t,y)=0. \quad (25.33)$$

当(25.33)成立时，(25.31)和(25.32)可写为在 S_δ 上积分的截尾矩形式.

仿一维情形可定义 n 维扩散过程. 在$\frac{\partial F}{\partial x_i}$与$\frac{\partial^2 F}{\partial x_i \partial x_j}$存在且连续的假定下，可建立 n 维扩散过程的科尔莫戈罗夫向后方程

$$\frac{\partial F}{\partial s}=-\sum_{i=1}^{n}a_i(s,x)\frac{\partial F}{\partial x_i}-\frac{1}{2}\sum_{i,j=1}^{n}b_{ij}(s,x)\frac{\partial^2 F}{\partial x_i \partial x_j}. \quad (25.34)$$

如 n 维转移密度函数 $p(s,x;t,y)$存在，则可以 p 易(25.34)之 F. 类似地向前方程是

$$\frac{\partial p}{\partial t}=-\sum_{i=1}^{n}\frac{\partial[a_i(t,y)p]}{\partial y_i}+\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2[b_{ij}(t,y)p]}{\partial y_i \partial y_j}. \quad (25.35)$$

将一维扩散过程的讨论平行移植到 n 维扩散过程，无实质性困难. 详细的讨论见专著[78]，其中还以现代方法一鞅，作为工具.

26 布朗运动

26.1 引言与定义

26.1.1 引言与例

布朗运动过程(Brown motion processes),有时称为**维纳(Wiener)过程**,是应用概率论中最有用的随机过程之一.英国植物学家布朗(Robert Brown)首先观察到了,一个完全没入液体中或气体里的小微粒是怎样作不规则运动的. 1905 年爱因斯坦(Einstein)首先解释了这种现象,认为是该粒子不断地受到周围介质分子的碰撞造成的.维纳在 1918 年给出了布朗运动的定义.

布朗运动有效地用在品行的统计、检验,分析股票市场的价格水平,量子力学,以及系统控制等方面.布朗运动的研究带动随机分析、鞅论及随机微分方程的研究;它与古典位势间联系的发现,导致对马尔可夫过程与位势理论之间联系的深刻研究,从而成为现代概率论的一个重要生长点.

例 26.1.1 布朗运动 布朗运动可以解释为对称的**随机游动**(random walk)的极限.考虑节 25.1.2 中引例,X_t 是一粒子在时刻 $t(\geqslant 0)$的位置.在式(25.2)中令 $C=0$,则 $p=q=1/2$,为对称的游动.令 $\Delta t\to 0$,作极限过渡,知$\{X_t,t\geqslant 0\}$是平稳独立增量过程,且 $X_t\sim N(0,2Dt)$,$D>0$.当 $D=\frac{1}{2}$,此过程即为布朗运动过程(25.3).

26.1.2 d 维布朗运动定义

定义 26.1.2 **d 维布朗运动** 概率空间$(\Omega,\mathscr{F},P)$上的 d 维

连续随机过程 $B=\{B_t,t\geqslant 0\}$，且满足：

(1) 对任 n 及 $0\leqslant t_0<t_1<\cdots<t_n$，增量 $B_{t_1}-B_{t_0},B_{t_2}-B_{t_1},\cdots,B_{t_n}-B_{t_{n-1}}$ 相互独立；

(2) 对任 $s\geqslant 0,t>0,B_{t+s}-B_s\sim N(0,tI)$，即均值向量为 0，协方差阵为 tI 的 d 维正态分布，这里 I 为 $d\times d$ 单位阵.

$B_0=0$ 时的布朗运动，称为**标准的**(standard).

注 26.1.3 记号 BM＝布朗运动

由定义知，$B_{s+t}-B_s$ 的概率密度为

$$p(t,x)=(2\pi t)^{-d/2}\exp(-|x|^2/2t),\quad x\in R^d. \quad (26.1)$$

B_t 的分布则依赖于开始分布，即 B_0 的分布 $\mu(\cdot)=P(B_0\in\cdot),\cdot\in\mathscr{B}^d$. 由 B_0 与 B_t-B_0 独立及卷积公式，知对 $A\in\mathscr{B}^d$

$$P(B_t\in A)=\int_A\left[\int_{R^d}(2\pi t)^{-d/2}\exp(-|x-y|^2/2t)\mu(\mathrm{d}x)\right]\mathrm{d}y. \quad (26.2)$$

若令 $\widetilde{B}_t=B_t-B_0$，则 $\widetilde{B}=\{\widetilde{B}_t,t\geqslant 0\}$ 也是 BM，且有零初值. 因此常可设 BM 有零初值，或径直将零初值写进定义中去，此时它是一马尔可夫过程.

26.2 布朗运动的性质

性质 26.2.1 不变性(invariance)

(1) **正交**(orthogonal)**不变性** 设 H 是 R^d 中正交变换，则 $HB\triangleq\{HB_t,t\geqslant 0\}$ 也是 BM；

(2) **平移**(translation)**不变性** $a+B\triangleq\{a+B_t,t\geqslant 0\}$ 也是 BM，其中 $a\in R^d$；

(3) **尺度**(scale)**不变性** $\{B(ct)/\sqrt{c},t\geqslant 0\}$，仍是 BM，其中 $c>0$.

性质 26.2.2 列维(Lévy)振动性质 对$\forall\ s_1=t_0<t_1<\cdots<t_n=s_2$, $\Delta t_k=t_{k+1}-t_k$, $h=\max\Delta t_k$, $\Delta B_{t_k}=B_{t_{k+1}}-B_{t_k}$,有

$$E\left|\sum_0^{n-1}(\Delta B_{t_k})^2-(s_2-s_1)\right|^2\leqslant 2h(s_2-s_1).$$

性质 26.2.3 分析性质 BM 的样本函数以概率 1 连续,但在 t 的任何有限区间都不是有界变差的.

BM 这一特性使它常可用来描述市场及股票等价格的变化,在电信和控制系统中刻画噪声和随机扰动.这一性质还使 B_t 在普通意义下处处不可微.寻找新的微分定义并研究它们,成为经典随机分析一主要内容,大大促进了鞅论与随机微分方程论的发展,并使它们在应用领域成为有力的工具(参见节 30.2,30.5,31.1 及 31.6 等).

定理 26.2.4 **最大值分布** 设 $d=1$ 且 $B=(B_t)$为标准 BM,则

$$P(\max_{s\leqslant t}B_s\geqslant a)=\frac{1}{\sqrt{2\pi t}}\int_a^{\infty}\mathrm{e}^{-x^2/2t}\mathrm{d}x$$

定理 26.2.5 **重对数律**(law of the iterated logarithm) 设 $d=1$,

$$\lim_{t\downarrow 0}\sup B_t/\sqrt{2t\log\log t^{-1}}=1$$

定理 26.2.6 **反正弦律**(arc sine law) 设 $B=(B_t)$为一维标准 BM,则对任 $x\in(0,1)$,

$$\begin{aligned}P(\text{区间}(xt,t)\text{中 }B\text{ 无零})&=\frac{2}{\pi}\arcsin\sqrt{x}\\&=P([0,t]\text{中 }B\text{ 为正的总时间}\leqslant xt).\end{aligned}$$

下面讨论 d 维 BM 的转移概率密度.设 $t>0$, $x,y\in R^d$,令

$$p(t,x,y)=p(t,y-x)=(2\pi t)^{-d/2}\exp(-|y-x|^2/2t).\tag{26.3}$$

当开始分布集中在点 x 上,即 $\mu(\{x\})=P(B_0=x)=1$,则由(26.2),

$$P_x(B_t\in A)=\int_A p(t,x,y)\mathrm{d}y.\tag{26.4}$$

这里以 P_x 易 P 是强调开始分布的作用.

由此式,可解释 $p(t,x,y)$为:自点 x 出发作 BM 的粒子,于时刻 t 转移到点 y 附近的转移概率密度.

性质 26.2.7 转移概率密度的性质

(1) $p(t,x,y)$ 对 x,y 对称,且是**空间齐次的**(spatial homogeneous).

(2)

$$g(x,y)\triangleq\int_0^\infty p(t,x,y)\mathrm{d}t$$

$$=\begin{cases}\Gamma\left(\dfrac{d}{2}-1\right)/(2\pi)^{d/2}|x-y|^{d-2}, & \text{当 } d\geqslant 3,\\ \infty, & d\leqslant 2.\end{cases}$$

(3) 科尔莫戈罗夫向前方程为

$$\frac{\partial}{\partial t}p(t,x,y)=\frac{1}{2}\frac{\partial^2}{\partial y^2}p(t,x,y).$$

向后方程为

$$\frac{\partial}{\partial t}F(t,x,y)=\frac{1}{2}\frac{\partial^2}{\partial x^2}F(t,x,y).$$

对有界 $\mathscr{B}^d$ 可测函数 f,定义变换 T_t:

$$T_tf(x)=\int_{R^d}f(y)p(t,x,y)\mathrm{d}y,\quad t>0.$$ 则

(4) $\{T_t\}$为压缩算子半群.

(5) $\{T_t\}$的强无穷小算子 $\mathrm{A}=\Delta/2$,Δ 为拉普拉斯算子. 即对有界、二次连续可微、二阶偏导有界且在 R^d 上均匀连续的函数 f,有

$$\mathrm{A}f(x)=\frac{1}{2}\sum_{i=1}^d\frac{\partial^2f(x)}{\partial x_i^2},\quad x=(x_1,x_2,\cdots,x_d).\quad(26.5)$$

性质 26.2.8 作为马尔可夫过程的 BM

(1) 如 B 为零初值的 BM,即标准 BM,则由定义 26.1.2 知 B

为马尔可夫过程. 又若令 $\mathscr{N}_t^s=\sigma\{B_u,s\leqslant u\leqslant t\}$(即在时间段$[s,t]$上 BM 产生的 σ 代数), $\mathscr{N}_\infty^0\triangleq\bigcup\limits_{t\geqslant 0}\mathscr{N}_t^0$,则对固定 a, $B^a\triangleq B+a$(注意定义 26.2.1(2))在 $\mathscr{N}_\infty^a$ 上可产生满足下列条件的唯一概率测度 P_a: 对任意的 $A_i\in\mathscr{B}^d$, $i=\overline{1,n}$, 及 $0\leqslant t_1<\cdots<t_n$,

$$
\begin{aligned}
&P_a(B_{t_1}^a\in A_1,\cdots,B_{t_n}^a\in A_n)\\
=&P(B_{t_i}+a\in A_i,i=\overline{1,n})\\
=&\int_{A_1}p(t_1,a,\mathrm{d}a_1)\int_{A_2}p(t_2-t_1,a_1,\mathrm{d}a_2)\cdots\\
&\int_{A_n}p(t_n-t_{n-1},a_{n-1},\mathrm{d}a_n).
\end{aligned}
$$

则$\{B^a,a\in R^d\}$构成一马尔可夫过程$\{B_t^a,\mathscr{N}_t^0,P_a,t\geqslant 0,a\in R^d\}$,简记为$(B_t,\mathscr{N}_t,P_x)$,其转移密度为(26.3). 它是由各点出发的 BM 所共同组成.

(2) $(B_t,\mathscr{N}_t,P_x)$及$(B_t,\mathscr{N}_{t-},P_x)$都是强马尔可夫过程,这里 $\mathscr{N}_{t+}=\bigcap\limits_{s>t}\mathscr{N}_s^0$.

BM 的等价定义与推广:

定理 26.2.9 对$(\Omega,\mathscr{F},P)$上零初值的连续过程 B,下述命题等价:

(1) B 是 BM;

(2) 对$\forall\ \lambda\in R^d$ 及 $t>s\geqslant 0$,

$$E(\mathrm{e}^{\mathrm{i}\lambda^\tau(B_t-B_s)}|\mathscr{N}_s^0)=\mathrm{e}^{-\frac{1}{2}|\lambda|^2(t-s)}\quad[P];$$

(3) 对$\forall\ \lambda\in R^d$ 及 $t>s\geqslant 0$,

$$E(\mathrm{e}^{\mathrm{i}\lambda^\tau(B_t-B_s)}|\mathscr{N}_{s-})=\mathrm{e}^{-\frac{1}{2}|\lambda|^2(t-s)}\quad[P].$$

由此可将 BM 定义稍作推广:以一般的参考族$(\mathscr{F}_t)$代替由过程生成的参考族$(\mathscr{N}_t^0)$. 所谓**参考族**(reference family)指子 σ 代数$\{\mathscr{F}_t\}(\subset\mathscr{F})$,且满足 $\mathscr{F}_t\subset\mathscr{F}_s$, $t<s$ 及 $\mathscr{F}_{t+0}=\mathscr{F}_t$.

定义 26.2.10 广义 BM 或($\mathscr{F}_t$)BM 对($\Omega,\mathscr{F},P$)上 d 维连续的、$\{\mathscr{F}_t\}$适应过程 $B=\{B_t,t\geqslant 0\}$，如满足：$\forall\ \lambda\in R^d$ 及 $t>s\geqslant 0$，有

$$E\left(e^{i\lambda^T(B_t-B_s)}\mid\mathscr{F}_s\right)=e^{-\frac{1}{2}|\lambda|^2(t-s)}\qquad[P].$$

注意此时 B_0 未必是零，因此 B 未必是马尔可夫过程. 但可证 B_t-B_s 与 $\mathscr{F}_s$(从而与 $\mathscr{F}_0$ 可测的 B_0)独立. 又如(B_t)是($\mathscr{F}_t$)BM，则(B_t-B_0)也是，新过程(B_t-B_0)是马尔可夫过程.

BM 的其他定义 BM 也可定义为满足 $EX_t=0,\mathrm{cov}(X_{s+t},X_t)=tI(s\geqslant 0)$的 Gauss 过程[71,187]. 零初值 BM 还可由扩散过程获得定义(节 25.3.4)，漂移系数为 0，扩散系数为 1.

26.3 首中时与常返性

26.3.1 首中时与首中点

近代马尔可夫过程论中一个极重要的概念是首中某集的时间. 对 d 维 BM，集 $A\in\mathscr{B}^d$ 的**首中时**(hitting time)h_A 定义为

$$h_A=\begin{cases}\inf(t:\geqslant 0,B_t\in A),\\ \infty,\text{如上述 } t \text{ 集为空},\end{cases}$$

它是第一次击中(到达)A 的闭包的时刻，也是 $A^c=R^d\backslash A$ 的首出时.

1. 首中时性质

性质 26.3.1 h_A 为马尔可夫时间(24.3.4).

性质 26.3.2 **零一律**(zero or one law) $P_x(h_A=0)=0$ 或 $1,\forall\ x\in R^d$.

定义 26.3.3 如 $P_x(h_A=0)=1$，称 $x\in R^d$ 为 A 的**规则点**(regular point)；否则称为 A 的**非规则点**(irregular point).

A 的规则点全体记为 A^r. A 的边界 $\partial A\triangleq\bar{A}\backslash A^0$ 上的点既可是

规则点，也可是非规则点，这里 $\overline{A}$ 为 A 之闭包，A^0 为 A 之内点全体.

性质 26.3.4 $A^0 \subset A^r \subset \overline{A}$，而 $L(A \backslash A^r)=0$，其中 $L(\cdot)$表勒贝格(Lebesque)测度.

性质 26.3.5 $f(x,t) \triangleq P_x(h_A \leqslant t)$，$x \in R^d$，$t>0$ 则

(1) 固定 x，$f(x,t)$是 t 的连续函数；固定 t，对 x 它是**下半连续的**(lower semi-continuous)，即$\varliminf\limits_{y \to x} f(y,t) \geqslant f(x,t) > -\infty$.

(2) $\lim\limits_{t \to \infty} P.(t \leqslant h_A < \infty)=0$ 在紧集上为一致.

(3) A 为紧集时 $f(x,t)$是热传导问题

$$\begin{cases} \dfrac{\partial f}{\partial t} = \dfrac{1}{2}\Delta f, \quad t>0, \quad x \in A^c \triangleq R^d \backslash A, \\ f(x,0)=0, x \in A^c \text{ 及} \lim\limits_{x \to y} f(x,t)=1, y \in A \cap A^r \end{cases} \tag{26.6}$$

的唯一解.

据此可视 $f(x,t)$为时刻 t 在 $x \in A^c$ 上的温度，而 $\int_{A^c} f(x,t)\mathrm{d}x$ 为时刻 t 自 A 流入周围介质 A^c 中的总能量 $E_A(t)$. $d=3$，$t \to \infty$时 $E_A(t)=tC(A)+4(2\pi)^{-3/2}[C(A)]^2 t^{1/2}+O(t^{1/2})$，其中 $C(A)$为 $A \in \mathscr{B}^d$的容度(26.5.11).

性质 26.3.6 对 h_A 用首次通过公式，有

(1) 设 $A, A_1 \in \mathscr{B}^d$，则

$$P_x(B_t \in A_1) = P_x(B_t \in A_1, h_A > t) + \int_0^t \int_{\overline{A}} P_a(B_{t-s} \in A_1) P_x(h_A \in \mathrm{d}s, B(h_A) \in \mathrm{d}a);$$

(2) $$E_x f(B_t) = E_x(f(B_t), h_A > t) + \int_0^t \int_{\overline{A}} E_a f(B_{t-s}) P_x(h_A \in \mathrm{d}s, B(h_A) \in \mathrm{d}a),$$

其中 $B(h_A)$为首中点，$f(x)$为任一可积函数.

2．球面首中时与首中点性质

记 h_r 为 d 维球面 $S_r=\{x:|x|=r\}, r>0$ 的首中时，这里 $|x| = \left(\sum^{d} x_i^2\right)^{1/2}, d\geqslant 2$.

定理 26.3.7 (1) $P_a(h_r<\infty)=1$，如 $|a|\leqslant r$；$=\left(\frac{r}{|a|}\right)^{d-2}$，如 $|a|>r$.

(2) $E_0 h_r=r^2/d$；

(3) $E_a h_r=(r^2-|a|^2)/d,\ |a|\leqslant r, d\geqslant 2$.

(4) $P_0(h_r>t)=\sum_1^{\infty}\zeta_{di}\exp\left(-\frac{q_{di}^2}{2r^2}t\right), t\geqslant 0$,

其中 q_{di} 是贝塞尔函数 $J_\nu(z), \nu=\frac{d}{2}-1$ 的正零点，又 $\zeta_{di}=q_{di}^{\nu-1}/2^{\nu-1}\Gamma(\nu+1)J_{\nu+1}(q_{di})$.

(5) $E_a\mathrm{e}^{-\lambda h_r}=\left(\frac{r}{|a|}\right)^{\nu} I_\nu(\sqrt{2\lambda}|a|)/I_\nu(\sqrt{2\lambda}r)$,

$$E_0\mathrm{e}^{-\lambda h_r}=(r\sqrt{2\lambda})^{\nu}/[2^{\nu}I_\nu(r\sqrt{2\lambda})\Gamma(\nu+1)].$$

其中 I_ν 为修正的贝塞尔函数，$\nu=\frac{d}{2}-1, d\geqslant 2, 0<|a|<r, \lambda>0$.

定理 26.3.8 (1) $P_0(B(h_r)\in A)=\mathrm{U}_r(A)$，可测集 $A\subset S_r$，即对 P_0，$B(h_r)$有球面上的均匀分布.

(2) $P_x(B(h_r)\in A)=\int_A r^{d-2}\left||x|^2-r^2\right||a-x|^{-d}\,\mathrm{U}_r(\mathrm{d}a)$，

$\forall$ 球内的点 x，$|x|<r$.

3．一维情形

$d=1$ 时 BM 的首中时，有下述公式

$$P_x(h_x=0)=1,$$

$$P_y(h_x\in\mathrm{d}t)=|y-x|(2\pi t^3)^{-\frac{1}{2}}\mathrm{e}^{-(y-x)^2/2t}\mathrm{d}t,$$

这里 h_x 为首中点 x 的时间.

26.3.2 常返性

定义 26.3.9 如 $P_x(L_A=\infty)=1,\forall x\in R^d$,则称集 A 为**常返的**(recurrent),否则为**暂留的**(transient),这里 L_A 为 A 的**末离时**(last exit time or quitting time),即

$L_A=\sup\{t:0<t,X_t\in A\};=0$,如前述 t 集空.

这里$\{X_t,t\geqslant 0\}$为任一随机过程.又如 $P_x(\lim\limits_{t\to\infty}|X_t|=\infty)=0,\forall x\in R^d$,则称此过程常返;如上一概率为 1,称过程暂留.

性质 26.3.10 BM 的常返性

(1) 如 $A\in\mathscr{B}^d$ 且 $P.(h_A<\infty)=1$ 在 R^d 上成立,则 A 常返,否则为暂留集.

(2) $d\leqslant 2$ 时非极集为常返的,$d\geqslant 3$ 时任一有界集为暂留集.

所谓**极集**(polar set),指使 $P.(h_A<\infty)=0$ 在 R^d 上成立的集 A. $d=1$ 时,任一非空集为常返,故非极集.而 $d\geqslant 2$ 时,任一单点集为极集;相对紧集[紧集(compact)]为极集的充要条件是其容度(见定义 26.5.11)为 0[其为疏集].

(3) BM 在 $d\leqslant 2$ 时常返,$d\geqslant 3$ 时暂留.

26.4 布朗运动的变种与布朗桥

26.4.1 布朗运动变种及应用

设 B 为 $d=1$ 的 BM.这里介绍几类 B 的函数的过程,它们实际上已不是 BM.以 $W_t=\sigma B_t+\mu t$,即有漂移系数 μ、扩散系数为 σ 的布朗运动(定义 26.4.10)来代替下述定义中的 B_t,从而拓广概念和讨论.

定义 26.4.1 设 h_x 为$\{x\}$的首中时,令

$$Z_t=B_t,\text{如 } t<h_x;\quad =x,\text{如 } t\geqslant h_x.$$

称$\{Z_t, t\geqslant 0\}$是在 x **有吸引的** BM(Brownian motion absorbed at x),它的分布有离散与连续两部分:

$$\begin{cases} P(Z_t = x) = \dfrac{2}{\sqrt{2\pi t}}\displaystyle\int_x^{\infty} \exp(-y^2/2t)\mathrm{d}y, \\ P(Z_t \leqslant y) = P(B_t \leqslant y) - P(B_t \leqslant y, \max\limits_{s\leqslant t} B(s) > x) \\ \qquad = \dfrac{1}{\sqrt{2\pi t}}\displaystyle\int_{y-2x}^{y} \exp(-u^2/2t)\mathrm{d}u, \quad y < x. \end{cases}$$

定义 26.4.2 令 $Z_t = |B_t|, t\geqslant 0$,称 $Z_T = \{Z_t, t\geqslant 0\}$为**在原点反射**(reflected at the origin)**的** BM,其 d.f 及矩分别为

$$\begin{cases} 2/\sqrt{2\pi t}\displaystyle\int_{-\infty}^{y} \exp(-x^2/2t)\mathrm{d}x - 1, \quad y > 0; \\ EZ_t = \sqrt{2t}/\pi, \quad V(Z_t) = (1-2/\pi)t. \end{cases}$$

反射 BM 的 Скороход 分解,可读[38,325].

定义 26.4.3 称 $Y_t = \mathrm{e}^{B_t}, t\geqslant 0$ 为**几何**(geometric)BM.

设 $d=1, B_0=0$,由 $E\mathrm{e}^{\lambda B_t} = \mathrm{e}^{t\lambda^2/2}$得

$$EY_t = \mathrm{e}^{t/2}, V(Y_t) = \mathrm{e}^{2t} - \mathrm{e}^t.$$

例 26.4.4 几何 BM 应用 几何 BM 是很有用的模型,它适用于某些随机量比值变化为独立同分布情形.例如 Y_n 表时刻 n 某商品价格,并认为 $Y_n/Y_{n-1} \triangleq X_n$ 为 i.i.d(即独立同分布),且 $Y_0=1$,则 $Y_n = X_1X_2\cdots X_n$.对 $\log Y_n$ 作适当规范化,由中心极限定理知可近似为 BM,从而$\{Y_n\}$为渐近几何 BM.

例 26.4.5 设某股票现在价格为 y,其价格变化 Y 为几何 BM.此股票在将来某时刻 T 可自所购公司兑得固定的价值 $k(>y)$.某人买了这份股票,到 T 时可期望得到更高收益的期望是

$$E\max[(Y_T - k), 0] = \frac{1}{\sqrt{2\pi T}}\int_0^{\infty}\int_{\log[(k+a)/y]}^{\infty} \mathrm{e}^{-x^2/2T}\mathrm{d}x\mathrm{d}a.$$

定义 26.4.6 称 $Z_t = \int_0^t B_s \mathrm{d}s, t \geqslant 0$ 为**积分** BM(integrated Brownian motion).

结论 26.4.7 基本结论

(1) $\{Z_t, t\geqslant 0\}$是高斯过程,但不是马尔可夫过程;

(2) $EZ_t=0, \mathrm{cov}(Z_s, Z_t)=s^2\left(\frac{t}{2}-\frac{s}{6}\right), \quad s\leqslant t$;

(3) $\{(B_t, Z_t), t\geqslant 0\}$是马尔可夫过程;

(4) $(B_t, Z_t)\sim N(0,0,t,t^3/3,\sqrt{3}/2)$.

例 26.4.8 设某商品价格变化率就是当时的通货膨胀率,按 BM 变化,则以 Z_t 表示 t 时刻价格时,$\frac{\mathrm{d}Z_t}{\mathrm{d}t} = B_t$,从而 $Z_t = Z_0 + \int_0^t B_s \mathrm{d}s$. $\{Z_t - Z_0, t \geqslant 0\}$ 即积分 BM.

例 26.4.9 如此商品价格 W_t 变化率不按 BM 变化,而满足 $\frac{\mathrm{d}}{\mathrm{d}t}W_t=B_tW_t$. 取 $W_0=1$ 则 $W_t=\mathrm{e}^{Z_t}$,这里$\{Z_t, t\geqslant 0\}$为积分 BM. 此时 $EW_t=\mathrm{e}^{t^6/6}$.

定义 26.4.10 称过程 $X=\{X_t, t\geqslant 0\}$是**有漂移系数为 μ 的** BM(BM with drift μ),如满足

(1) $X_0=0, \quad X_t\sim N(\mu t, t)$,

(2) X 有平稳独立增量.

换言之,$X_t=B_t+\mu t$,B 为零初值 BM. X 有速率为 μ 的一个漂移趋势. 若令 $W_t=\sigma B_t+\mu t$,则称为有漂移系数为 μ、扩散系数为 σ^2 的 BM.

作为随机游动极限的有漂 BM 设一随机游动向右一步概率为 $p = \frac{1}{2}(1 + \mu\sqrt{\Delta t})$,向左为 $1-p$,且在(25.2)中取 $\Delta x= \sqrt{\Delta t}$,$\Delta t\to 0$即得有漂移系数 μ 的 BM. 也可作为扩散过程的特例 可视为漂移系数为 μ 扩散系数为 1 的扩散过程(见定义

25.2.1). 此时尺度函数 $S(x)=(1-e^{-2\mu x})/(2\mu)$, 速度密度 $m(x)=e^{2\mu x}$.

性质 26.4.11　性质与公式

(1) $\max\limits_{0\leqslant s\leqslant t} X_s/t \xrightarrow[t\to\infty]{} \mu$, [P]如 $\mu\geqslant 0$.

(2) 设 $\lambda>0, x>0$, 首中时的拉普拉斯变式

$$Ee^{-\lambda h_x}=\begin{cases}\exp\{x(\mu+\sqrt{\mu^2+2\lambda})\} & \mu<0,\\ \exp\{-x(\sqrt{\mu^2+2\lambda}-\mu)\}, & \mu\geqslant 0.\end{cases}$$

(3) 设 $b>0, -b<x<a$,

$$u(x)\triangleq P_x(h_a<h_{-b})=(e^{2\mu b}-e^{-2\mu x})/(e^{2\mu b}-e^{-2\mu a}).$$

进一步的性质可由扩散过程性质得到. 应用的例子可读[71]节 6.4.

26.4.2　布朗桥

定义 26.4.12　布朗桥(Brownian bridge)定义为均值为 0 协方差函数为 $s(1-t/T_0)$, $0\leqslant s\leqslant t<T_0$ 的高斯过程, $X\triangleq\{X_t, 0\leqslant t\leqslant T_0\}$.

由定义 26.4.12 知 $X_0=X_{T_0}=0$, a.e, 故过程轨道以概率 1 两端系定形成一"桥".

性质 26.4.13　基本结论

(1) 如 B 为 BM, 则 $X_t=B_t-B_0-\frac{t}{T_0}(B_{T_0}-B_0)$, $0\leqslant t\leqslant T_0$ 为布朗桥, 且它与 $\{B_0, B_{T_0}\}$ 独立.

(2) $X_t=(t-T_0)\int_0^t (s-T_0)^{-1}\mathrm{d}B_s$ 是布朗桥.

(3) $\mathrm{d}X_t=\frac{X_t}{t-T_0}\mathrm{d}t+\mathrm{d}B_t$ 的在 $0\leqslant t\leqslant T_0$ 中零初值的解是布朗桥.

(4) 条件随机过程 $\{B_t, 0\leqslant t\leqslant 1 \mid B_1=0\}$ 是一[0,1]上的布朗桥, 这里 B_t 为 BM.

(5) 布朗桥 X 的有限维分布函数为

$$P(X_{t_i} < x_i, i = \overline{1,n})$$
$$= (p(T_0,0,0))^{-1}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} p(t_1,0,y_1)p(t_2,-t_1,y_1,y_2)\cdots$$
$$p(T_0-t,y_n,0)\mathrm{d}y_1\cdots\mathrm{d}y_n, 0<t_1<\cdots<t_n<T_0,$$

$p(t,x,y)$为 BM 的转移密度.

(6) 布朗桥 X 是马尔可夫过程,且可逆,即 $X_{T_0-t}, 0\leqslant t\leqslant T_0$ 也是布朗桥.

(7) X 是$[0,T_0]$上布朗桥,则 $X_t/(T_0-t)$是$[0,T_0)$上独立增量过程,且为高斯鞅,其特征(见式 29.7)为 $t/T_0(T_0-t)$.

(8) 布朗桥与经验分布函数　设$\{X_n\}$i.i.d,$N_n(s)$为 $X_1,X_2,\cdots,X_n$ 中不超过 $s\in(0,1)$的个数,则 $F_n(s)=N_n(s)/n$ 为经验 d.f. (14.2).这时

(i) 如 $X_i\sim U_{(0,1)}$,则 $\alpha_n(s)\triangleq\sqrt{n}(F_n(s)-s), s\in[0,1]$的极限过程($n\to\infty$)是$[0,1]$上布朗桥;

(ii) 如 X_i 的 d.f $F(x)$连续,则 $\alpha_n(s)\triangleq\sqrt{n}(F_n(y_s)-F(y_s))$ 的极限过程是$[0,1]$上布朗桥 Z,这里 $y_s=F^{-1}(s)$,且

$$\lim_{n\to\infty}P\left\{\sqrt{n}\sup_x|F_n(x)-F(x)|<a\right\}=P\{\max_{0\leqslant t\leqslant 1}|Z_t|<a\}.$$

26.5　高维布朗运动的牛顿势

设 $d\geqslant 3$,B 为 BM.

26.5.1　调和函数

定义 26.5.1　调和函数(harmonic function)　开集 $A\subset R^d$ 中实函数 $h(x)$,在 A 中连续,$\dfrac{\partial^2 h}{\partial x_i^2}$存在且满足拉普拉斯方程:$\Delta h\triangleq\sum\limits^{d}\dfrac{\partial^2 h}{\partial x_i^2}=0$.

定理 26.5.2 邓肯定理 设 A 为相对紧(即 $\overline{A}$ 为紧集)开集，u 在 $\overline{A}$ 连续，Δu 在 A 中存在连续且有界，则对一切 $a\in\overline{A}$，有

$$E_a[u(B_e)]-u(a)=\frac{1}{2}E_a\int_0^e\Delta u(B_s)\mathrm{d}s, \qquad (26.7)$$

其中 e 为 A 的首出时.

定理 26.5.3 函数 $h(x)$ 在开集 A 中调和的充要条件是它在 A 中**局部可积**(locally integrable)(即在开集 A 中每一紧集上 L 可积)且有**球面平均性**，即对 $\forall\ a\in A$ 及球面

$$S_r(a)=(x;|x-a|=r)\subset A, r>0,$$

$$h(a)=\int_{S_r(a)}h(x)U_r(\mathrm{d}x)=E_0h(B_{e_r}) \qquad (26.8)$$

其中 e_r 为 $S_r(a)$ 的首中时，也是开球 $\mathring{B}_r(a)\triangleq(x;|x-a|<r)$ 的首出时.

当 h 调和时，还可证明有球体平均性，利用这一性质可证 R^d 上调和函数若有上界或下界，则必为常数.

定理 26.5.4 **极值原理**(extremum principle)或极大极小值原理(principle of maximum and minimum values)

在有界开集 A 中调和且在 $\overline{A}$ 连续的函数 h，有

$$\inf_{x\in\partial A}h(x)\leqslant h(a)\leqslant\sup_{x\in\partial A}h(x), \forall\ a\in\overline{A},$$

其中 ∂A 指 A 的边界点全体.

26.5.2 狄利克雷问题

狄利克雷(Dirichlet)问题(简称 D 问题)，是要求一在开集 $A(\subset R^d, d\geqslant 2)$ 中调和、在 $\overline{A}$ 连续的函数 h，使其在边界 ∂A 上等于预先给定的连续函数 f. 此问题是 1840 年高斯提出，1924 年维纳提出广义 D 问题. 角谷静夫(Kakutani)(1944 年)和杜勃(Doob)(1954 年)发现它与 BM 间的联系，用概率方法对 D 问题研究的主要结论是

定理 26.5.5 设 A 为有界开集，则 D 问题有解的充要条件是 $\partial A \subset (A^c)^r$. 此时解为唯一，且可表示为 $h(x)=E_x f(B_e)$，$x \in \bar{A}$.

A 无界时，解可有无穷多个. 当开集 A 未必有界，但 f 在 ∂A 不仅连续且有界，则 D 问题有解为

$$h(x) = E_x[f(B_e), e < \infty] + cP_x(e = \infty), \quad c = \text{const}. \tag{26.9}$$

定理 26.5.6 庞加莱锥判别法(Poincare cone criterion)如存在以 $x \in \partial A$ 为顶点的锥 $K=(b: \in R^d, |(b-x)u| \geqslant a|b-x|)$ 及球 $B_r(x)=(b: \in R^d, |b-x| \leqslant r)$，使 $K \cap B_r(x) \subset A$，则 $x \in A^r$，这里 u 为单位向量，a 为某常数.

26.5.3 测度的势与扫除问题

定义 26.5.7 测度 μ 的**牛顿势**(Newton potential)

$$G_\mu(x) = \int_{R^d} g(y - x)\mu(\mathrm{d}y), d \geqslant 3,$$

势核 $g(y-x) \triangleq g(x,y) = \Gamma\left(\frac{d}{2}-1\right)|x-y|^{2-d}/2\pi^{d/2}$.

定理 26.5.8 (1) 如 $G_\mu(x)<\infty$，则

$$G_\mu(x) - T_t G_\mu(x) = \int_0^t T_s \mu(x)\mathrm{d}s;$$

(2) μ 有限时，G_μ 唯一决定 μ.

(3) **极大值原理**(maximum principle) 设 μ 有限，支集为 A (A 是一切使 $\mu(U)=0$ 的开集 U 的和的补集)，且 $N \subset A$，$\mu(N)=0$，如 $G_\mu(\cdot) \leqslant M < \infty$ 在 $N^c \cap A$ 上成立，则 $\sup\limits_{x \in R^d} G_\mu(x) \leqslant M$.

公式 26.5.9 势的基本公式(fundamental identity for potential) $g(y-x) = \int_{\bar{A}} H_A(x, \mathrm{d}a) g(y-a) + g_A(x,y)$，

其中 $H_A(x,\cdot) = P_x(B(h_A) \in \cdot)$，即首中点分布，

$$g_A(x,y)=\int_0^\infty q_A(t,x,y)\mathrm{d}t,$$

而 $$q_A(t,x,y)\mathrm{d}y=P_x(h_A>t,B_t\in\mathrm{d}y).$$

例 26.5.10 扫除(balayage)问题 已给 $A\in\mathscr{B}^d$ 及 $\mu\in\mathscr{U}=\{$有限测度 ν:$G_\nu(\cdot)$ 为局部可积$\}$. 试求测度 $\mu'\in\mathscr{U}$,其**支集**(support)含于 A^r,且使 $G_{\mu'}(x)\leqslant G_\mu(x)$,$x\in R^d$;在 A^r 上等号成立.

扫除问题的解 设 A 为紧集,则该问题有唯一解

$$\mu':\mu'(\cdot)=\int H_A(x,\cdot)\mu(\mathrm{d}x).$$

26.5.4 平衡测度

设 A 为相对紧集,则测度 $\mu_A(\mathrm{d}y)\triangleq \mathrm{s}\lim\limits_{|x|\to\infty}H_A(x,\mathrm{d}y)/g(x)$ 存在且对球面 S_r,$\mathring{B}_r\supset\overline{A}$,有

$$\mu_A(\mathrm{d}y)=\int_{R^d}\frac{r^{d-2}}{2\pi^{d/2}}\Gamma\left(\frac{d}{2}-1\right)U_r(\mathrm{d}x)H_A(x,\mathrm{d}y),$$

其中 s lim 指测度的强收敛,即 $\mathrm{s}\lim\nu_n=\nu$ 指 $\sup\limits_{A\in\mathscr{B}^d}|\nu_n(A)-\nu(A)|\to 0,n\to\infty$.

定义 26.5.11 称上述测度 μ_A 为 A 的**平衡测度**(equilibrium measure). 而此测度的**全质量**称为 A 的**容度**(capacity),记 $C(A)$.

$\forall\ \Gamma\in\mathscr{B}^d$ 有

$$\lim_{|x|\to\infty}P_x(B(h_A)\in\Gamma|h_A<\infty)=\mu_A(\Gamma)/C(A),\tag{26.10}$$

故规范化后的平衡测度可理解为自"∞"出发,A 的首中点的条件分布.

平衡测度 μ_A 的势 G_{μ_A} 称为**平衡势**(equilibrium potential). 当 A 相对紧时,$G_{\mu_A}(x)=P_x(h_A<\infty)$,$x\in R^d$;对球面 S_r,其平衡测度及容度

$$\mu_{S_r}(\mathrm{d}y)=2\pi^{d/2}r^{d-2}U_r(\mathrm{d}y)/\Gamma\left(\frac{d}{2}-1\right); \qquad (26.11)$$

$$C(S_r)=\frac{d-2}{2r}|S_r|. \qquad (26.12)$$

性质 26.5.12 容度的性质

(1) 单调性：如 $A\subset\Gamma$，则 $C(A)\leqslant C(\Gamma)$.

(2) $C(A+x)=C(A)$，这里 $A+x=(a+x:a\in A)$.

(3) $C(\alpha A)=\alpha^{d-2}C(A)$，常数 $\alpha>0$.

(4) 如 A 为相对紧开集，则

$$C(A)=\sup\{C(K):K\subset A,K\text{ 为紧集}\}.$$

(5) 如 A 为紧集，则 $C(A)=\inf\{C(O):$开集 $O\supset A$，且 $\overline{O}$ 为紧集$\}$.

26.6 一维二维布朗运动的位势

对一维和二维 BM，$G_\mu(x)=\infty$(26.5.7)，故需另找势核，借助拉普拉斯变换，发现 $d=2$ 时势核 $k(x,y)=\frac{1}{\pi}\log|y-x|$，$d=1$ 时 $k(x,y)=|y-x|^{-1}$. 由此分别建立**对数位势**(logarithmic potential)与**线性位势**(linear potential)，仿牛顿位势作平行讨论，略述如下(相应记号不变).

26.6.1 平面的对数位势

对数势的基本公式 设 A 非极集，$x\neq y$，$k(y-x)=k(x,y)$，则 $k(y-x)=\int_A H_A(x,\mathrm{d}a)k(y-a)-g_A(x,y)+L_A(x)$，

(26.13)

其中 $$L_A(x)\triangleq\lim_{\lambda\downarrow 0}(1-E_x\mathrm{e}^{-\lambda\tau_A})\int_0^\infty \mathrm{e}^{-\lambda t}p(t,x)\mathrm{d}t.$$

性质 26.6.1 $L_A(x)$的性质

(1) 如 A 为相对紧且非极集，则

$$\lim_{|y|\to\infty} g_A(x,y) = L_A(x), \qquad \lim_{|x|\to\infty} g_A(x,y) = L_A(y).$$

(2) 如 $E_x h_A<\infty$（例如 A^c 相对紧时即可）则 $L_A(x)\equiv 0$.

(3) 如 A 为相对紧，则存在有限极限 $\lim\limits_{|x|\to\infty}[L_A(x) - k(x)] \triangleq R(A)$. 又若测度 $\mu_A(\mathrm{d}y) \triangleq \int_{S_r} H_A(x,\mathrm{d}y)U_r(\mathrm{d}x)$ 的势（定义 26.6.2）为 K_{μ_A}，则 $L_A(x) = R(A) - K_{\mu_A}(x)$，这里 S_r 为圆周.

(4) 设 A 为有界非极集，则

$$L_A(x) = \lim_{t\to\infty} \frac{1}{2\pi}\log t P_x(h_A > t), x \in R^2.$$

圆的情形　设 $B_r=\{b:\in R^2, |b|\leqslant r\}$，$S_r$ 为圆周，

(1) $L_{B_r}(x)=\frac{1}{\pi}\log(|x|/r)I_{(B_r)^c}(x)$；

(2) $R(B_r)=\pi^{-1}\log(1/r)$；

(3) $E_x h_{S_r}=\frac{1}{2}(r^2-|x|^2), x\in B_r$

定义 26.6.2　有界测度 ν 如有紧支集 C，令 $K\nu(x) = -\int_C k(y - x)\nu(\mathrm{d}y)$，称 K_ν 为 **v 的势**.

如 A 相对紧且非极集，则称(26.6.1)中 μ_A 为 A 的**平衡测度**，K_{μ_A} 为**平衡势**，$R(A)$ 为 Robin **常数**，且可证 $\mu_A(\mathrm{d}y) = \mathrm{s}\lim\limits_{|x|\to\infty} H_A(x,\mathrm{d}y)$.

26.6.2　直线上的线性位势

线性位势的基本公式

$$k(x,y) = \int_{\bar{A}} H_A(x,\mathrm{d}a)k(a,y) - g_A(x,y) + L_A(x). \tag{26.14}$$

$g_A(x,y)$计算公式　设 A 非空，$(a,b)\subset(\overline{A})^c$，$x\in[a,b]$，则

(1) $x\in[a,b]$时，$g_A(x,y)=0$，如 $y\notin(a,b)$；
$g_A(x,y)=[(b-x)(y-a)+(x-a)(b-y)]/(b-a)-|y-x|$，如 $y\in(a,b)$.

(2) $+\infty>x>b_0\triangleq\sup(a:a\in A)$时
$g_A(x,y)=x+y-2b_0-|y-x|$，如 $y>b_0$；$=0$，如 $y\leqslant b_0$.
$-\infty<x<a_0\triangleq\inf(a:a\in A)$时，

$g_A(x,y)=2a_0-x-y-|y-x|$，如 $y<a_0$；$=0$，如 $y\geqslant a_0$.

其他结果

(1) 如 A 有界非空，则 $\mu_A=(\delta_{a_0}+\delta_{b_0})/2$，

$$R(A)=(a_0-b_0)/2.$$

(2) $L_A(x)=\lim\limits_{t\to\infty}(\pi t/2)^{1/2}P_x(h_A>t)$，　$x\in R_1$.

27 独立增量过程与更新过程

27.1 定义和一般性质

本章考虑定义在某概率空间$(\Omega,\mathscr{F},P)$上，以$T=[0,+\infty)$为参数集，R^1为状态空间的随机过程. 其结果可推广到R^n和T为左闭有穷或无穷区间的情形.

定义 27.1.1 若对任意$n,t_k\in T,t_1<t_2<\cdots<t_n$，随机变量$X_0,X_{t_1}-X_0,X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}}$是相互独立的，则随机过程$X=\{X_t\}_{t\in T}$称为**独立增量过程**(processes with independent increment).

独立增量过程可视为独立随机变数和序列在连续时间情形的推广. 此类过程的有限维分布由X_0的分布及所有增量$X_t-X_s(s\leqslant t,s,t\in T)$分布$F_{st}$决定. 下设$X_0=0$.

定义 27.1.2 若对所有$s,t\in T,s\leqslant t,X_t-X_s$的分布$F_{st}$仅与$t-s$有关，则称独立增量过程$\{X_t\}_{t\in T}$为齐次的(homogeneous).

独立增量过程是马尔可夫过程，其转移函数由F_{st}表出. 对齐次独立增量过程，有下面的结果.

定理 27.1.3 设$\{X_t\}_{t\in T}$有齐次独立增量，那么作为马尔可夫过程其转移函数$P(t,x,\Gamma)$满足：

$$P(t,x_0+x,\Gamma+x_0)=P(t,x,\Gamma),x_0,x\in R^1,\quad(27.1)$$

其中$\Gamma\in\mathscr{B},\Gamma+x_0\triangleq\Gamma+\{x_0\}$；反之转移函数满足(27.1)式的马尔可夫过程为独立增量过程.

由此可知齐次独立增量过程是**费勒过程**.

注 独立增量过程也有以$X_{t_1}-X_0,X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}}$相

互独立来定义的，此时零初值（即 $P(X_0=0)=1$）的独立增量过程才是马尔可夫过程．一般地，如令 $\tilde{X}_t=X_t-X_0$，则过程 $\tilde{X}_{t,t\in T}$，可保证为零初值的，从而仿建上述结论．

定义 27.1.4 随机连续的独立增量过程称为**列维过程**（Lévy process）．

设 $\{X_t\}_{t\in T}$为列维过程，则增量 X_t-X_s 的分布 F_{st}是无穷可分的，因此它的特征函数 φ_{st}具有形式：

$$\log\varphi_{st}(\theta)=\mathrm{i}\theta(m(t)-m(s))+\int_{R^1}\left(\mathrm{e}^{\mathrm{i}\theta x}-1-\frac{\mathrm{i}\theta x}{1+x^2}\right)\frac{1+x^2}{x^2}\mathrm{d}_x[G(t,x)-G(s,x)].$$

其中 m 为连续函数，$G(t,x)$关于 t 连续，关于 x 为有界单调，且 $G(t,\cdot)-G(s,\cdot)(s<t)$也是单调的．在齐次情形 $m(t)-m(s)$和 $G(t,\cdot)-G(s,\cdot)$都线性地依赖于 $t-s$．

通常可用特征函数刻画随机连续性．

定理 27.1.5 独立增量过程在 t_0 处随机连续的充要条件是 F_{0t}的特征函数 φ_{0t}在 t_0 处连续．

27.2 基本独立增量过程

27.2.1 泊松过程

定义 27.2.1 设$\{N_t\}_{t\in T}$是非负整数值过程，若存在$[0,+\infty)$上的非减实函数 $\lambda(\cdot)$，$\forall\, k=0,1,\cdots$，满足

$$P(N_t-N_s=k)=\frac{1}{k!}[\lambda(t)-\lambda(s)]^k\exp\{\lambda(s)-\lambda(t)\}. \tag{27.2}$$

则称为**泊松过程**（Poisson process）．若 $\lambda(t)=\lambda t$，则称$\{N_t\}_{t\in T}$为**齐次泊松过程**．

下面结果说明泊松过程是一类具有特殊跳跃轨道的独立增量

过程.

定理 27.2.2 齐次列维过程为可分泊松过程的充要条件是它的轨道以概率 1 为非减、跃度为 1 的阶梯函数.

因此,跳跃点和跃度是泊松过程的两个主要特征.

定理 27.2.3 设$\{N_t\}_{t\in T}$为齐次泊松过程,参数为 λ.

(1) $\forall\ 0\leqslant m\leqslant n, 0\leqslant s\leqslant u<t$,

$$P(N_u - N_s = m \mid N_t - N_s = n) = \binom{n}{m}\left(\frac{u-s}{t-s}\right)^m\left(1-\frac{u-s}{t-s}\right)^{n-m}.$$

(2) 记第 n 个跳跃点为 η_n, $\tau_n=\eta_n-\eta_{n-1}$,则$\{\tau_n\}_{n\geqslant 1}$为独立同分布序列,且

$$P(\tau_n > t) = e^{-\lambda t}, n\geqslant 1, t\geqslant 0.$$

(3) 提供了一个构造泊松过程的方法:取一列独立同指数分布(参数为 λ)的随机变数列$\{\tau_n\}$,令

$$N_t = \inf\{n-1: \tau_1+\cdots+\tau_n > t\},\quad t\in T,$$

则$\{N_t\}_{t\in T}$为具有参数 λ 的齐次泊松过程.

27.2.2 由列维过程得到泊松过程

定理 27.2.4 可分列维过程没有第二类不连续点. 现设$\{X_t\}_{t\in T}$为可分列维过程,那么对任意 $\varepsilon>0$,以 $\mathscr{B}_\varepsilon$ 表示 $R^1/[-\varepsilon,\varepsilon]$ 上的博雷尔集类全体,显然它是一个 σ 代数(特称博雷尔代数);则对每一 $A\in\mathscr{B}_\varepsilon$, $t>0$,过程在$(0,t)$内具有跃度 $X_s - X_{s^-}\in A$ 的跳跃点数 $\nu(t,A)$以概率 1 是有限的,于是得到一族随机过程$\{\nu(t,A)\}_{t\in T}$, $A\in\mathscr{B}_\varepsilon$.

下一定理主要给出由列维过程得到一个泊松过程.

定理 27.2.5 (1) $E(\nu(t,A))^\gamma<\infty, \forall\ t\in T, A\in\mathscr{B}_\varepsilon, \gamma>0$.

(2) 若 $A_i\in\mathscr{B}_\varepsilon, i=1,\cdots,n$,且 $A_iA_j=\Phi, i\neq j$,则 n 个过程$\{\nu(t,A_i)\}_{t\in T}, i=1,2,\cdots,n$ 是相互独立的.

(3) $\forall A \in \mathscr{B}_\varepsilon$，$\{\nu(t,A)\}_{t\in T}$是泊松过程，相应的参数函数$\lambda(t) = E(\nu(t,A)) = \Pi(t,A)$，它是连续增的，且$\lim\limits_{\varepsilon\to 0}\int\limits_{\varepsilon<|x|<1} x^2\Pi(x,\mathrm{d}x) < \infty$，$t \in T$。

27.2.3 维纳过程

定义 27.2.6 一个增量 $X_t - X_s$ 具有正态分布 $N(m(t)-m(s), \nu(t)-\nu(s))$的独立增量过程称为维纳(Wiener)过程 $W(m(\cdot),\nu(\cdot))$. 这里 $m(\cdot)$和 $\nu(\cdot)$为 T 上的实函数，后者是增加的. 若 $m(\cdot)$和 $\nu(\cdot)$具有形式：$m(t)=mt$，$\nu(t)=\nu t$，m，ν 为常数，则称相应过程为**齐次维纳过程**.

关于齐次维纳过程(也叫布朗运动)可参阅第 26 章.

27.3 独立增量过程的分解

27.3.1 列维分解

设$\{X_t\}_{t\in T}$为独立增量过程，则对任意实函数 $f(t)$，$\{X_t - f(t)\}_{t\in T}$也是独立增量的，列维证明了，适当选择 f 可使 $X_t - f(t)$ 具有某种连续性.

定义 27.3.1 若过程 $X_t - f(t) = \xi_t$ 具有以下性质：

(1) $\forall t \geqslant 0$，ξ_s 在 $s=t$ 有左右极限 ξ_t^-，ξ_t^+；

(2) $\xi_t^+ - \xi_t$ 和 $\xi_t - \xi_t^-$ 之一为常数时二者皆为 0；

(3) 除至多可列个 $t \in T$ 外，以概率 1 有 $\xi_t^- = \xi_t = \xi_t^+$. 则实函数 $f(t)$，$t \in T$ 称为 X 的**中心化函数**. 此时称 ξ_t 为**中心化过程**(centralized process).

定理 27.3.2 每一个独立增量过程$\{X_t\}_{t\in T}$可以表为如下形式：

$$X_t = f(t) + \alpha(t) + \beta(t),$$

称为列维分解. 其中 $f(t)$ 为中心化函数，α,β 为中心化独立增量过程，除了至多可列个固定 t 值外 α 是连续的，而 β 没有第二类不连续点，且

$$E(\operatorname{arctg}\alpha(t)) = E(\operatorname{arctg}\beta(t)) = 0, t \in T.$$

27.3.2 伊藤分解

若 $\{X_t\}_{t\in T}$ 是可分列维过程，$\{\nu(t,A)\}_{t\in T}, A\in \mathscr{B}_\varepsilon$ 是由它得到的泊松过程族，记

$$\tilde{\nu}(t,A) = \nu(t,A) - \Pi(t,A),$$

则 X 可由维纳过程和泊松过程表示.

定理 27.3.3 伊藤(Ito)分解　对每个可分列维过程 X，存在一个维纳过程 $\xi=\{\xi_t\}_{t\in T}$，它独立于 $\{\nu(t;A)\}_{t\in T}$，$A \in \bigcup\limits_{\varepsilon>0}\mathscr{B}_\varepsilon$，使

$$X_t = \xi_t + \int\limits_{|x|>1} x\nu(t,\mathrm{d}x) + \int\limits_{|x|\leqslant 1} x\tilde{\nu}(t,\mathrm{d}x). \qquad (27.3)$$

上式称为 X 的伊藤分解.

由 X 的伊藤分解，可得出它的特征函数 φ_{st} 的另一种等价形式：

$$\varphi_{st}(\theta) = \exp\Big\{\mathrm{i}m(t) - \frac{\theta^2}{2}\nu(t) + \int\limits_{|x|>1}(\mathrm{e}^{\mathrm{i}\theta x} - 1)\Pi(t,\mathrm{d}x)$$

$$+ \int\limits_{0<|x|\leqslant 1}(\mathrm{e}^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta x)\Pi(t,\mathrm{d}x).$$

其中 $m(\cdot)$ 和 $\nu(\cdot)$ 是 ξ 的均值和方差函数，它们是连续的，ν 是单增的.

由(27.3)式可推知，若 X 的轨道是阶梯函数，则特征函数 φ_{0t} 具有下述形式：

$$\log\varphi_{0t}(\theta) = \int_R (\mathrm{e}^{\mathrm{i}\theta x} - 1)\Pi(t,\mathrm{d}x)$$

若 X 的轨道是增加的，则 $m(t)$ 是增加的，

又 $$\log\varphi_{0t}(\theta)=\mathrm{i}m(t)\theta+\int_0^{\infty}(\mathrm{e}^{\mathrm{i}\theta x}-1)\Pi(t,\mathrm{d}x)$$

且 $$\int_0^1 x\Pi(t,\mathrm{d}x)<\infty.$$

27.4 样本函数性质

设$\{X_t\}$为任一齐次列维过程，那么特征函数φ_{st}的对数$\log\varphi_{st}$对s,t的依赖是关于$t-s$的线性函数，确切地：

$$\log\varphi_{s,s+t}(\theta)=\log\varphi_t(\theta)=tk(\theta),$$

此处

$$k(\theta)=\mathrm{i}m\theta-\frac{\nu}{2}\theta^2+\int\limits_{|x|>1}(\mathrm{e}^{\mathrm{i}\theta x}-1)\Pi(\mathrm{d}x)$$
$$+\int\limits_{|x|\leqslant 1}(\mathrm{e}^{\mathrm{i}\theta x}-1-\mathrm{i}\theta x)\Pi(\mathrm{d}x),$$

m和ν为常数. 通称$k(\theta)$为X的**累积量**(cumulant).

先考虑X的样本函数的局部性质，由于$X_{t+s}-X_t$与X_s同分布，所以只需考虑过程在0点的行为.

定理 27.4.1 (1) 如果X的变差有界，而

$$k(\theta)=\mathrm{i}m\theta+\int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}\theta x}-1)\Pi(\mathrm{d}x).$$

则

$$P(\lim_{t\to 0}X_t/t=m)=1.$$

(2) 若X的变差无界，则以概率1有

$$\overline{\lim_{t\to 0}}X_t/t=+\infty,\underline{\lim_{t\to 0}}X_t/t=-\infty.$$

考虑单调过程的局部性质.

定理 27.4.2 设X为齐次独立增量过程，且

$$K(\theta)=\int_0^{\infty}(\mathrm{e}^{\mathrm{i}\theta x}-1)\Pi(\mathrm{d}x).$$

其次设 $g(x)$是定义在$[0,\infty)$上的函数，满足条件：$g(0)=0$，连续单增，且 $g(x+y)\leqslant g(x)+g(y)$，那么

(1) 如 $\int_0^\infty g(x)\Pi(\mathrm{d}x)<\infty$，则 $P(\lim_{t\to 0}g(X_t)/t=0)=1$.

(2) 若 $\int_0^\infty g(x)\Pi(\mathrm{d}x)=+\infty$，则 $P(\lim_{t\to 0}g(X_t)/t=+\infty)=1$.

下面定理比较细致地刻画了任意过程的局部增长.

定理 27.4.3 设$\{X_t\}_{t\in T}$是齐次独立增量过程，$\varphi(t)$是$[0,1]$上一连续的非负递增函数，满足下列条件：

(1) $\lim_{u\downarrow 1}\ \sup_t|\varphi(u,t)/\varphi(t)-1|=0$.

(2) 对$\forall\ \varepsilon>0$，存在 $\alpha_\varepsilon>0$，使 $P(X_t<-\varepsilon\varphi(t))\leqslant 1-\alpha_\varepsilon$.

那么 (1)如果 $\int_0^1 1/tP(X_t>\varphi(t))\mathrm{d}t<\infty$，则

$$P\{\overline{\lim_{t\to 0}}X_t/\varphi(t)\leqslant 1)=1.$$

(2) 如果$\int_0^1 1/tP(X_t>\varphi(t))\mathrm{d}t=\infty$，则

$$P(\overline{\lim_{t\to 0}}X_t/\varphi(t)\geqslant 1)=1$$

对一类特殊过程，可以求出 $\varphi(t)$. 熟知，X_t 为标准维纳过程($m=0,\nu=1$)，成立重对数定理 26.2.5. 若 X 为指数 $\alpha\in[1,2)$的稳定过程，且只有**负跳跃**(negatively jump)，即其累积量为

$$K(\theta)=-c|\theta|^\alpha\left(1-\mathrm{i}\frac{\theta}{|\theta|}\omega(\theta,\alpha)\right).$$

其中 c 为常数，而

$$\omega(\theta,\alpha)=\begin{cases}\mathrm{tg}\,\dfrac{\pi}{2}\alpha, & \alpha\neq 1,\\[2ex] \dfrac{2}{\pi}\ln|\theta|, & \alpha=1.\end{cases}$$

那么对应的增长函数 φ 亦可求出.

定理 27.4.4 如果 $X=\{X_t\}_{t\in T}$是只有非负跳跃的稳定过程，指数 $\alpha\in[1,2)$，则

$$P(\overline{\lim_{t\to 0}} X_t/\varphi(t) = 1) = 1.$$

其中

$$\varphi(t) = \begin{cases} \alpha(\alpha-1)^{\frac{1-\alpha}{\alpha}}(c_1 t)^{\frac{1}{\alpha}}\left[\ln\ln\frac{1}{t}\right]^{\frac{\alpha-1}{\alpha}}, & 1 < \alpha < 2, \\ \frac{2ct}{\pi}\ln\frac{1}{t}, & \alpha = 1 \end{cases}$$

$c_1 = c/\cos\frac{\pi}{2}\alpha.$

现考虑在无穷远处的增长.

定理 27.4.5 若对齐次独立增量过程 X 存在均值 EX_1(可为无穷),则

$$P\left(\lim_{t\to\infty}\frac{1}{t}X_t = EX_1\right) = 1.$$

注意此结果是强大数律在连续参数情形的推广. 若 $EX_1 = 0$, 下面结果说明 X 的振动仍具有无穷振幅.

定理 27.4.6 设 X 具有齐次独立增量, $EX_1 = 0$, 那么

$$P(\sup_t X_t = +\infty) = P(\inf_t X_t = -\infty) = 1$$

27.5 某些泛函的分布

设 $\{X_t\}_{t\in T}$ 是齐次独立增量过程,它的累积量为

$$K(\theta) = \mathrm{i}a\theta - \frac{b\theta^2}{2} + \int_{0<|x|\leqslant 1}(\mathrm{e}^{\mathrm{i}\theta x} - \mathrm{i}\theta x - 1)\Pi(\mathrm{d}x) + \int_{|x|\geqslant 1}(\mathrm{e}^{\mathrm{i}\theta x} - 1)\Pi(\mathrm{d}x)$$

1. 上、下确界的分布

令

$$\eta_t^+ = \sup_{0\leqslant s\leqslant t} X_s, \quad q^+(\lambda, x) = \lambda\int_0^\infty P(\eta_t^+ < x)\mathrm{e}^{-\lambda t}\mathrm{d}t$$

$$\eta_t^- = \inf_{0\leqslant s\leqslant t} X_s, \quad q^-(\lambda,x) = \lambda\int_0^\infty P(\eta_t^- < x)e^{-\lambda t}dt$$

定理 27.5.1 分布 $q^+(\lambda,x)$ 和 $q^-(\lambda,x)$ 的特征函数分别为：

$$\int e^{izx}d_xq^+(\lambda,x) = \exp\left\{\int_0^\infty \frac{1}{t}e^{-\lambda t}\int_0^\infty (e^{izx}-1)dF_t(x)dt\right\},$$

$$\int e^{izx}d_xq^-(\lambda,x) = \exp\left\{\int_0^\infty \frac{1}{t}e^{-\lambda t}\int_{-\infty}^0 (e^{izx}-1)dF_t(x)dt\right\},$$

其中 $F_t(x)=P(X_t\leqslant x)$.

推论 27.5.2 $q^+(\lambda,x)$ 和 $q^-(\lambda,x)$ 连续的充分必要条件分别是

$$\int_0^\infty \frac{1}{t}e^{-\lambda t}P(X_t > 0)dt = +\infty$$

和

$$\int_0^\infty \frac{1}{t}e^{-\lambda t}P(X_t < 0)dt = +\infty.$$

若 X 为阶梯过程，则 $q^+(\lambda,x)$ 和 $q^-(\lambda,x)$ 对一切 $x\neq 0$ 连续.

2. 首中时及首中点分布

对 $x\in R^1$，设 τ_x 为 $\{X_t\}$ 首中区间 $(x,+\infty)$ 的时刻，记 $\gamma_x = X_{\tau_x} - x$.

定理 27.5.3 随机变量 τ_x 和 γ_x 的联合拉普拉斯变换为

$$Ee^{-\lambda\tau_x-\mu\gamma_x} = 1 - q^+(\lambda,x) - \frac{\mu}{\lambda^2}\int_{0^-}^{x}\int_{-\infty}^{0^+}\int_0^\infty e^{-\mu y}M(x-s+y-a)dy\,d_sq^-(\lambda,s)d_\mu q^+(\lambda,\mu).$$

它们的联合分布为

$$P(\tau_x\langle t,\gamma_x\rangle y) = \int_{0^-}^{x}\int_{-\infty}^{0^+} M(x-s+y-u)R_t(ds,du)$$

其中 $M(x) = \int_x^\infty \Pi(dy)$,

$$R_t(A,B) = \int_0^t Q^+(B,t-s)ds\,Q^-(A,s),$$

而
$$Q^{+}(B,s)=P(\eta_s^{+}\in B),$$
$$Q^{-}(A,s)=P(\eta_s^{-}\in A)$$

3. 上、下确界与过程值的联合分布

注 27.5.4 符号
$$Q(t;a,b;\alpha,\beta)=P(\eta_t^{-}\geqslant a,\eta_t^{+}\leqslant b,\alpha<X_t<\beta),$$
其中 $a<0<b,a\leqslant\alpha<\beta\leqslant b$.
$$\Gamma^{+}(x,\mathrm{d}t,\mathrm{d}y)=P(\tau_x\in \mathrm{d}t,\gamma_x\in \mathrm{d}y),x>0,$$
$$\Gamma^{-}(x,\mathrm{d}t,\mathrm{d}y)=P(\tau_x'\in \mathrm{d}t,\gamma_x'\in \mathrm{d}y),x<0,$$
其中 τ_x' 是首达 $(-\infty,x)$ 的时间，$\gamma_x'=X_{\tau_x'+0}-x$.
$$\tilde{Q}(\lambda;a,b;\alpha,\beta)=\int_0^\infty \mathrm{e}^{-\lambda t}Q(t;a,b;\alpha,\beta)\mathrm{d}t.$$
$$\hat{\Gamma}_\lambda^{+}(x)=\int_0^\infty \mathrm{e}^{-\lambda t}\Gamma^{+}(x,\mathrm{d}t,[x,\infty)),$$
$$\hat{\Gamma}_\lambda^{-}(x)=\int_0^\infty \mathrm{e}^{-\lambda t}\Gamma^{-}(x,\mathrm{d}t,(-\infty,x]).$$
$$R_\lambda(A)=\int_0^\infty \mathrm{e}^{-\lambda t}P(X_t\in A)\mathrm{d}t.$$

定理 27.5.5 对 $\lambda\geqslant0,a<0<b,a\leqslant\alpha<\beta\leqslant b$，有
$$\begin{aligned}\tilde{Q}(\lambda;a,b;\alpha,\beta)&=R_\lambda((\alpha,\beta))\\&-\frac{\Gamma_\lambda^{+}(b)[R_\lambda(\alpha-b,\beta-b)-\Gamma_\lambda^{-}(a-b)R_\lambda(\alpha-a,\beta-a)]}{1-G_\lambda^{-}(b,\{b\})}\\&-\frac{\Gamma_\lambda^{-}(a)[R_\lambda(\alpha-a,\beta-a)-\Gamma_\lambda^{+}(b-a)R_\lambda(\alpha-b,\beta-b)]}{1-G_\lambda^{-}(a,\{a\})},\end{aligned}$$
其中
$$\begin{aligned}&G_\lambda^{-}(a,\{a\})\\&=\iint_0^\infty\int_0^\infty \mathrm{e}^{-\lambda(s+t)}\Gamma^{+}(b-2a,\mathrm{d}t,\mathrm{d}y)\Gamma^{-}(a-b-y,\mathrm{d}s,\{2a\}),\\&G_\lambda^{+}(b,\{b\})\\&=\iint_0^\infty\int_0^\infty \mathrm{e}^{-\lambda(s+t)}\Gamma^{-}(a-2b,\mathrm{d}t,\mathrm{d}y)\Gamma^{+}(b-a-y,\{0\}).\end{aligned}$$

27.6 广义泊松过程和复合泊松过程

本节介绍两种应用很广的独立增量过程，并给出一些例子.

27.6.1 广义泊松过程

定义 27.6.1 若齐次独立增量过程 $N=\{N_t\}_{t\in T}$ 取非负整数值且是增加的，则称它为**广义泊松过程**(generalized Poisson process).

广义泊松过程的特征函数具有简单的形式.

定理 27.6.2 设 $\{N_t\}_{t\in T}$ 为广义泊松过程，假定 $N_0=0$，$\forall\, t>0$，$0<P(N_t>0)<1$ 且存在数列 $\{\rho_k\}_{k\geqslant 1}$，使

$$\lim_{h\to 0} P(N_{t+h}-N_t=k\,|\,N_{t+h}-N_t\geqslant 1)=\rho_k,$$

则 N_t 的特征函数：

$$\varphi_t(\theta)=e^{\nu t(\varphi(\theta)-1)}, \tag{27.4}$$

其中 ν 为某常数，而 $\varphi(\theta)=\sum\limits_{k=1}^{\infty}\rho_k e^{ik\theta}$.

例 27.6.3 广义泊松过程的一个典型例子是一列泊松过程的线性组合：

设 $\{N_k(t)\}_{t\in T}$，$k=1,2,\cdots$ 为一列独立的齐次泊松过程，N_k 的参数(平均速率)为 λ_k，$k=1,2,\cdots$，假定 $\nu=\sum\limits_{k=1}^{\infty}\lambda_k<\infty$.

令
$$N_t=N_1(t)+2N_2(t)+\cdots+kN_k(t)+\cdots,$$

则 $\{N_t\}_{t\in T}$ 是一广义泊松过程，其特征函数 $\varphi_t(\theta)$ 由(27.4)给出，而 $\rho_k=\lambda_k/\nu$.

很多实际问题都可用广义泊松过程作为模型.

27.6.2 复合泊松过程

另一类应用得很多的独立增量过程是由一列独立随机变数和

一个泊松过程复合而成的过程.

定义 27.6.4 若随机过程$\{X_t\}_{t\in T}$可表为

$$X_t = \sum_{n=1}^{N_t} y_n, \quad t \in T, \tag{27.5}$$

其中$\{y_n\}_{n\geq 1}$为一列独立同分布的随机变量,$\{N_t\}_{t\in T}$是与$\{y_n\}_{n\geq 1}$独立的齐次泊松过程.则称它是复合泊松过程(compound Poisson proccess).

定理 27.6.5 一个复合泊松过程$\{X_t\}_{t\in T}$具有齐次独立增量,且对任意$t\geqslant 0$,X_t具有特征函数:

$$\varphi_t(\theta) = e^{\nu t(\varphi(\theta)-1)}, \tag{27.6}$$

其中$\varphi(\theta)$是y_1的特征函数,ν为某个常数.

例 27.6.6 对保险公司的总要求 设某人寿保险公司的保险单持有者在时刻$\tau_1,\tau_2,\cdots$死亡,其中$\tau_1<\tau_2<\cdots$,他们的死亡是强度为ν的泊松型事件.在时刻τ_n死亡者的保险金额为y_n,这些钱在他死亡时由保险公司付给.保险公司为了确定应当保持多少储备以便支付对它提出的赔偿要求,自然很想知道从0到t这段时间内将要支付的总金额X_t.显然X_t可以表示成(27.5)的形式,其中N_t为$(0,t)$内死亡人数,因此$\{X_t\}$为一复合泊松过程,利用(27.6)可以求出X_t有关特征量,比如:

$$EX_t = \nu t E y_1,$$
$$V_{ar} X_t = \nu t E y_1^2,$$

等等.于是通过对Ey_1的估计,求得保险公司的平均储备金额.

泊松过程的其它应用见(22.5)、第32、33章及[70]的第13、16章.

27.7 更新过程与更新函数

27.7.1 更新过程定义与例

定义 27.7.1 设 $X_1, X_2, \cdots$ 为独立非负同分布随机变量且 $F(0)=P(X_n=0)<1$，令 $S_n=X_1+X_2+\cdots+X_n$，记 $N(t)=\sup\{n: S_n\leqslant t\}$，则称 $N_T \triangleq \{N(t), t\geqslant 0\}$ 为**更新过程**(renewal process).

也有称$\{S_n, n=1,2,\cdots\}$为更新过程的. 它们之间有下述关系：

$$N(t)\geqslant n \Leftrightarrow S_n\leqslant t. \tag{27.7}$$

下面来看一下更新过程的实例：考虑灯泡的连续更换：一个灯泡在时刻 0 时开始使用，在时刻 X_1 时坏了，马上换上一个新灯泡；第二个灯泡在时刻 X_1+X_2 坏了，再换上第三个灯泡；以此类推，第 n 个灯泡在 $S_n=\sum\limits_{i=1}^{n}X_i$ 时烧坏，立刻换上第 $n+1$ 个灯泡，则 $N(t)$ 为直到时刻 t 需要替换的灯泡个数，它是一个**计数过程**(counting process).

例 27.7.2 泊松过程　设$\{N(t), t\geqslant 0\}$为具有参数 λ 的齐次泊松过程：

$$P(N_t-N_s=k)=\frac{\mathrm{e}^{-\lambda(t-s)}\lambda^k(t-s)^k}{k!}\quad \lambda>0,\quad k=0,1,2,\cdots$$

由(27.7)及 $S_n=\sum\limits_{i=1}^{n}X_i$，可推得：$\{X_n, n=1,2,\cdots\}$有相同分布：$F(x)=1-\mathrm{e}^{-\lambda x}, x\geqslant 0$，泊松过程$\{N(t), t\geqslant 0\}$是一特别的更新过程.

例 27.7.3 设 $\xi_1, \xi_2, \cdots$ 为实值独立同分布随机变量序列，假设 $E\{\xi_i\}\geqslant 0$，考虑：$\nu_0=0, \nu_n=\xi_1+\xi_2+\cdots+\xi_n, n\geqslant 1$. 令 $S_1=\inf\{n: \nu_n>0\}, S_k=\inf\{n: \nu_n>U_{S_{k-1}}\}, X_1=S_1, X_k=S_k-S_{k-1}, k=2,3,\cdots$ 则 $X_1, X_2, \cdots$ 为正的独立同分布整值随机变量，而 $S_m=X_1+X_2+\cdots$

$+X_m, m\geqslant 1$ 按 27.7.1 式定义的过程$\{N(t), t>0\}$为更新过程.

27.7.2 更新过程分布与更新函数

由(27.7)知,$P(N(t)=n)=P(N(t)\geqslant n)-P(N(t)\geqslant n+1)$ $=P(S_n\leqslant t)-P(S_{n+1}\leqslant t)$

因随机变量 $X_i, i\geqslant 1$ 为独立的,且有同分布 F,记 F_n 为 F 的 n 次卷积,则 $N(t)$的分布为

$$P(N(t)=n)=F_n(t)-F_{n+1}(t). \tag{27.8}$$

性质 27.7.4 更新过程的渐近性质

$$\lim_{t\to\infty}N(t)=\infty, \text{a.e.} \text{但}$$

$$\lim_{t\to\infty}N(t)/t=\frac{1}{\mu}, \quad \mu=EX_n.$$

定义 27.7.5 $m(t)\triangleq E[N(t)]$称为过程 N_T 的**更新函数**(renewal function). 若 $N(t)$具有泊松分布,$P(N(t)=k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}, k\geqslant 0$. 则 $m(t)=E[N(t)]=\lambda t$.

27.8 基本更新定理与更新过程

定理 27.8.1 基本更新定理(elementary renewal theorem) 设 N_T 为更新过程,$EX_1=\cdots=EX_n=\mu$,则

$$\lim m(t)/t=1/\mu \quad (\text{理解 } 1/\infty=0).$$

定义 27.8.2 设 $F(t)$为一正随机变量的分布函数,$a(t)$为已知函数,关于 $A(t)$的方程:

$$A(t)=a(t)+\int_0^t A(t-x)\mathrm{d}F(x), \quad t\geqslant 0$$

称为**更新方程**(renewal equation).

当 $a(t)$为更新过程中 X_n 的分布函数 $F(t)$时,更新函数 $m(t)$

是更新方程的解.对于一般的 $a(t)$,更新方程的解可由更新函数表出(见节 27.9).

定理 27.8.3 设 $a(t)$ 为一有界函数,则存在惟一的在任一有穷区间上有界的函数 $A(t)$,满足更新方程,且其解是

$$A(t) = a(t) + \int_0^t a(t - x)\mathrm{d}m(x), \tag{27.9}$$

其中 $m(t) = \sum_1^\infty F_k(t)$ 为更新函数,$F_k(t)$ 为 $F(t)$ 的 k 重卷积.

27.9 主要更新定理及其应用

27.9.1 主要更新定理

首先,给出算术函数和直接黎曼可积定义.

定义 27.9.1 设分布函数 F,若对每个 $\varepsilon>0$:$F(\alpha+\varepsilon)-F(\alpha-\varepsilon)>0$,则称 F 在点 α 为递增的.

若存在正数 λ 使 F 在 $0,\pm\lambda,\pm 2\lambda,\cdots$ 为递增点.分布函数 F 称为**算术的**(arithmatic),最大的 λ 称为 F 的**跨度**(span).显然,若 F 为连续函数,则 F 为非算术的.而有值 $0,1,2,\cdots$ 的离散随机变量的分布函数为具有跨度 1 的算术函数.

定义 27.9.2 设 g 为 $[0,\infty)$ 上函数,对每个 $\delta>0,n=1,2,\cdots$,令

$$\underline{m}_n = \min\{g(t):(n-1)\delta \leqslant t \leqslant n\delta\},$$

$$\overline{m}_n = \max\{g(t):(n-1)\delta \leqslant t \leqslant n\delta\},$$

$$\underline{\sigma}(\delta) = \delta\sum_{n=1}^{\infty}\underline{m}_n,\ \overline{\sigma}(\delta) = \delta\sum_{n=1}^{\infty}\overline{m}_n,$$

若对每个 $\delta>0$,$\underline{\sigma}(\delta)$ 和 $\overline{\sigma}(\delta)$ 绝对收敛,且 $|\overline{\sigma}(\delta)-\underline{\sigma}(\delta)|\to 0,\delta\to 0$,则称 g 为**直接黎曼可积**(directly Riemann integrable).当 g 为直接黎曼可积的单调函数,则称它是**绝对可积**(absolutely integrable).

显然，所有满足$\int_0^{\infty}|g(t)|\mathrm{d}t<\infty$的单调函数的有限组合为**直接黎曼可积的**.

定理 27.9.3 主要更新定理(basic renewal theorem) 设 F 为具有均值 μ 的正随机变量的分布函数，设 $a(t)$ 为一已知的直接的黎曼可积函数而 $A(t)$ 为**更新方程**(定义 27.8.2)的解：

(1) 若 F 为非算术的，则

$$\lim_{t\to\infty}A(t)=\begin{cases}\dfrac{1}{\mu}\displaystyle\int_0^{\infty}a(x)\mathrm{d}x, & 若\ \mu<\infty,\\ 0, & 若\ \mu=\infty.\end{cases}$$

(2) 若 F 为算术的，且具有跨度 λ，则对所有 $c>0$，

$$\lim_{n\to\infty}A(c+n\lambda)=\begin{cases}\dfrac{\lambda}{\mu}\displaystyle\sum_{n=0}^{\infty}a(c+n\lambda), & 若\ \mu<\infty,\\ 0, & 若\ \mu=\infty.\end{cases}$$

主要更新定理有下面等价形式，它用更新函数来刻画更为直接.

定理 27.9.4 关键更新定理(key renewal theorem) 设 F 为具有均值 μ 的正随机变量的分布函数，设 $m(t)=\sum_{k=1}^{\infty}F_k(t)$ 为更新函数，其中 F_k 为 F 的 k 次卷积，设 $h>0$ 为固定的.

(1) 若 F 为非算术的，则 $\lim_{t\to\infty}[m(t+h)-m(t)]=\dfrac{h}{\mu}$.

(2) 若 F 为算术的，h 为跨度 λ 的倍数，则仍有：$\lim_{t\to\infty}[m(t+h)-m(t)]=\dfrac{h}{\mu}$，$\mu=\infty$时，理解为$\dfrac{h}{\mu}=0$.

上述两定理证明见[71].

27.9.2 主要更新定理的应用

首先介绍几个概念

$Y(t)=S_{N(t)+1}-t$，称为在时间 t 的**剩余寿命**(excess life / residual life).

$A(t)=t-S_{N(t)}$，称为**年龄**(ages)或**现在寿命**(current life).

$I(t)=Y(t)+A(t)$，称为**全寿命**(total life). 现利用主要更新定理求剩余寿命 $Y(t)$ 的极限分布：

记 $A_z(t)=P\{Y(t)>z\}$，对每个固定 $z>0$，由 $Y(t)$ 的定义，有

$$P(Y(t)>z|X_1=x)=\begin{cases}1, & 若\ x>t+z,\\ 0, & 若\ t+z\geqslant x>t,\\ A_z(t-x), & 若\ t\geqslant x>0.\end{cases}$$

则由全概率公式：

$$A_z(t)=\int_0^{\infty}P(Y(t)>z|X_1=x)\mathrm{d}F(x),$$

$$=1-F(t+z)+\int_0^t A_z(t-x)\mathrm{d}F(x),$$

为得极限分布，假设 $\mu=E(X_1)=\int_0^{\infty}\{1-F(x)\}\mathrm{d}x<\infty$，则

$$\int_0^{\infty}\{1-F(t+z)\}\mathrm{d}t=\int_z^{\infty}\{1-F(y)\}\mathrm{d}y<\infty.$$

故当 z 固定时，$\{1-F(t+z)\}$ 作为 t 的单调函数，为直接黎曼可积的.

由(27.9.3)，剩余寿命 Y_t 的极限分布为

$$\lim_{t\to\infty}P(Y_t>z)=\lim_{t\to\infty}A_z(t)=\frac{1}{\mu}\int_z^{\infty}\{1-F(y)\}\mathrm{d}y, z>0.$$

$A(t)$ 与 $Y(t)$ 的示意如图 27.1.

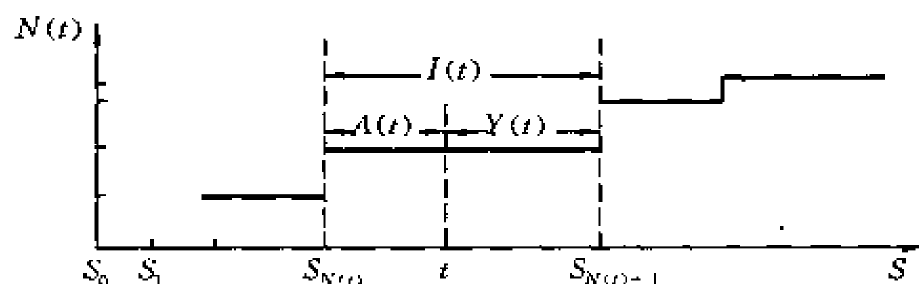


图 27.1

由上图，有

$$\{Y(t)\geqslant x, A(t)\geqslant y\}\Leftrightarrow\{Y(t-y)\geqslant x+y\}.$$

从而得极限联合分布：

$$\lim_{t\to\infty}P(Y(t)\geqslant x,A(t)\geqslant y)=\lim_{t\to\infty}P\{Y(t-y)\geqslant x+y\}$$

$$=\frac{1}{\mu}\int_{x+y}^{\infty}\{1-F(z)\}\mathrm{d}z.$$

特别地，令 $x=0$，得年龄 $A(t)$ 的极限分布：

$$\lim_{t\to\infty}P(A(t)\geqslant y)=\frac{1}{\mu}\int_{y}^{\infty}\{1-F(z)\}\mathrm{d}z.$$

最后，考虑 $I(t)=Y(t)+A(t)$ 的极限分布定义：$k_x(t)=P\{I(t)>x\}$，则

$$P\{I(t)>x\mid X_1=y\}=\begin{cases}1, & \text{若 } y>\max(x,t),\\ k_x(t-y), & \text{若 } y\leqslant t,\\ 0, & \text{其它}.\end{cases}$$

由全概率公式有：

$$k_x(t)=1-F(\max(x,t))+\int_0^t k_x(t-y)\mathrm{d}F(y)$$

由(27.9.3)，

$$\lim_{t\to\infty}P(I(t)>x)=\lim_{t\to\infty}k_x(t)=\frac{1}{\mu}\int_0^{\infty}[1-F(\max(x,\tau))]\mathrm{d}\tau$$

故，全寿命 $I(t)$ 的极限分布为

$$\lim_{t\to\infty}P(I(t)>x)=\frac{1}{\mu}\int_x^{\infty}\xi F(\xi),\quad x\geqslant 0.$$

28 分支过程

28.1 一维离散分支过程

记 $N=\{0,1,\cdots\}$，$N^{+}=\{1,2,\cdots\}$.

定义 28.1.1 设 $\{P_k\}_{k\in N}$ 为 N 上的一概率分布，$\{P_k^{*i}\}_{k\in N}$ 表示它的 i 重卷积. 如果它的转移函数 $P(i,j)$ 满足：

$$
\begin{aligned}
P(i,j) &= P(Z_{n+1}=j|Z_n=i) \\
&= \begin{cases} P_j^{*i}, i\geqslant 1, j\geqslant 0, \\ \delta_{0j}, i=0, j\geqslant 0. \end{cases}
\end{aligned}
$$

称以 N 为状态空间的马尔可夫链 $\{Z_n\}_{n\in N}$ 为**一维离散分支过程** (one dimensional discrete time branching process).

分支过程描述了下述模型：考虑一个群体（或系统），开始时有 Z_0 个个体，在一个单位时间后，每个个体按概率律 $\{P_k\}$ 彼此独立时群体数目，称它们为第一代. 然后经一单位时间后，每个第一代个体又独立地以 $\{P_k\}$ 产生第二代个体，如此下去，一般以 Z_n 表示第 n 代群体数目，则 $\{Z_n\}$ 就是一个分支过程，这里考虑群体由一种类型个体构造. 分支过程广泛应用于人口增长、细菌繁殖、链氏反应及计算机网络等. 离散一维分支过程亦称为**高尔顿-瓦特森** (Galton-Watson)过程，下面将简称为 GW **过程**.

作为一个马尔可夫链，GW 过程的状态空间具有较简单的结构：O 为吸引态，而 N^{+} 中状态都是非常返（暂留）的. 考虑其实际背景，对分支过程主要集中在两个问题上：群体消亡的概率及世代数（或时间充分大后的变化趋向）.

28.1.1 母函数及其方程

由于GW过程(后面将看到对一般分支过程亦如此)转移概率的特殊结构,使用转移概率的母函数 $\sum_{j\in N} P(i,j)s^j(|s|\leqslant 1)$ 代替转移概率本身是更方便的,而转移概率由 $\{P_k\}$ 决定,因此只需研究后者的母函数.定义GW过程 $\{Z_n\}_{n\in N}$ 的母函数为:

$$f(s)=\sum_{k\in N} P_k s^k,\ |s|\leqslant 1. \tag{28.1}$$

GW过程的各阶(步)转移概率的母函数可由 $f(s)$ 如下表出(设 $Z_0=1$):

记 $$f_n(s)=\sum_{j=0}^{\infty} P_n(1,j)s^j,\ n\geqslant 1,\ |s|\leqslant 1,$$

则 $$\sum_{j=0}^{\infty} P_n(i,j)s^j=[f_n(s)]^i,\ n\geqslant 1,$$

其中 $P_n(i,j)$ 是 n 阶(步)转移概率,注意 $f_1(s)=f(s)$,约定 $f_0(s)=s$,则有以下**母函数方程**(generating function equation):

$$f_n(f_l(s))=f_{n+l}(s),\quad n,l\in N,\ |s|\leqslant 1. \tag{28.2}$$

分支过程的大部分结果都是用母函数叙述,上述的半群关系(28.2)是很重要的.

前述 $Z_0=1$ 的假定并不失一般性,若开始有 i 个个体,则第 n 代数目的母函数(即转移概率 $P_n(i,j)$ 的母函数)为 $[f_n(s)]^i$,因此以下不特别说明时,总设 $Z_0=1$,此外为避免平凡情况还设 $P_0+P_1<1,P_j\neq 1,j\in N$.

28.1.2 矩和分类

假设第一代群体数 Z_1 的均值 m 和方差 σ^2 存在(即有限),那么 $m=f'(1)$,此时第 n 代数目 Z_n 的均值 $EZ_n=m^n$,方差

$$\mathrm{var}Z_n = \begin{cases} \dfrac{\sigma^2 m^{n-1}(m^n - 1)}{m-1}, & m \neq 1, \\ n\sigma^2, & m = 1. \end{cases}$$

m 和 σ^2 不但决定第 n 代的均值、方差，而且直接影响 Z_n 在 $n\to\infty$ 时的趋向，下面的分类将是这种影响的区分.

定义 28.1.2 如果相应的均值 m 分别为：>1，$=1$，<1，则称 GW 过程为**超临界**（supercritical）、**临界**（critical）或**次临界**（subcritical）.

28.1.3 灭绝概率

记 $q = p(Z_n = 0$，对某个 $n\in N^+)$ 它是群体最终消失（**灭绝**（extinction））的概率，下面定理给出 q 的解.

定理 28.1.3 设 $\{Z_n\}_{n\in N}$ 为 GW 过程，$m = EZ_1$，q 如上记号. 则当 $m\leqslant 1$ 时，$q=1$，当 $m>1$ 时，$q<1$，它是方程 $f(t)=t$ 的最小非负根.

注意：由于 $j\in N^+$ 都是非常返的，因此由(28.1.3)有

$$P(\lim_{n\to\infty} Z_n = 0) = 1 - P(\lim_{n\to\infty} Z_n = +\infty) = q$$

$$\lim_{n\to\infty} P(Z_n = j) = 0, \quad j \in N^+$$

28.1.4 极限性质

考虑 GW 过程 $\{Z_n\}_{n\in N}$ 的极限情形.

定理 28.1.4 若 $m<\infty$，则 $\{W_n = Z_n/m^n\}_{n\in N}$ 是一非负鞅，以 W 表 $\{W_n\}_{n\in N}$，当 $n\to\infty$ 时的 a.s. 极限，当 $m>1$ 时下列陈述等价：

(1) $\sum\limits_{k=2}^{\infty} KP_k \log K < \infty$，

(2) $P(W > 0) > 0$，

(3) $E|W - W_n| \to 0 \quad (n\to\infty)$，

(4) $\{W_n\}_{n\in N}$ 均匀可积，

(5) $E\sup_{n\in N}W_n<\infty$.

下面是更精细的结果.

1. 次临界情形

定理 28.1.5 若 $0<m<1$,则

$$\frac{f_n(s)-f_n(0)}{1-f_n(0)}\to g(s),n\to\infty,|s|\leqslant 1,$$

其中 g 是下列方程的唯一的概率母函数(probability generating function)解:

$$1-g(f(s))=m(1-g(s)),g(0)=0.$$

由此得出$\{Z_n\}$的条件极限分布:

定理 28.1.6 设 $0<m<1$,则存在概率分布$\{Q_k\}_{k\in N}$使

$$P(Z_n=j|Z_n>0)\to Q_j(n\to\infty).$$

2. 临界情形

定理 28.1.7 若 $m=1,\sigma^2<\infty$,则

$$\lim_{n\to\infty}\frac{1}{n}\left[\frac{1}{1-f_n(t)}-\frac{1}{1-t}\right]=\frac{\sigma^2}{2},$$

对 $0\leqslant t<1$ 均匀地成立.

用过程$\{Z_n\}_{n\in N}$来描述,有

定理 28.1.8 若 $m=1,\sigma^2<\infty$,则

$$\lim_{n\to\infty}P(Z_n/n>z|Z_n>0)=\exp(-2z/\sigma^2),z\geqslant 0.$$

3. 超临界情形

定理 28.1.9 若 $1<m<\infty$,则存在常数列$\{C_n\}_{n\in N}$满足 $C_{n+1}/C_n\to m$,使 $W_n=C_n^{-1}Z_n$ a.s. 收敛于具有以下性质的随机变量 W:

(1) $P(W>0)=1-q$.

(2) $\varphi(z)=E(e^{-zW})$满足

$$\varphi(z)=f\left(\varphi\left(\frac{z}{m}\right)\right).$$

(3) 若 $q=0$，W 具有在 $(0,\infty)$ 上绝对连续的分布.

28.2 一维连续马尔可夫分支过程

在(28.1)研究的模型中，其他假设不变，考虑每个个体不再是在单位时间内产生后代，而是随机生存一段时间 τ，该个体消失同时产生后代，个体间寿命是独立的且产生的后代数亦独立于其寿命，以 Z_t 表示时刻 t 群体的数目，则得到连续时间分支过程 $\{Z_t\}_{t\geqslant 0}$，仍设只有一种类型个体，故此过程是一维的.

假定 τ 服从参数为 a 的指数分布，个体产生的后代数分布仍为 $\{P_k\}_{k\in N}$，那么可以证明，$\{Z_t\}_{t\in N}$ 是一个离散参数的齐次马尔可夫过程，其密度矩阵 $Q=(q_{ij})_{i,j\in N}$ 为：

$$q_{ij}=\begin{cases}0, & j<i-1,\\ -ia, & i=j,\\ iaP_{j-i+1}, & j\geqslant i-1, j\neq i\end{cases}$$

以 $P_{ij}(t)$ 表示相应的最小 Q 过程，则：

$$\sum_{j=0}^{\infty}P_{ij}(t)s^j=\left[\sum_{j=0}^{\infty}P_{ij}(t)s^j\right]^i, i\in N, |s|\leqslant 1, \quad (28.3)$$

称之为分支条件. 一般给出下述

定义 28.2.1 概率空间 $(\Omega,\mathscr{F},P)$ 上的马尔可夫过程 $\{Z_t\}_{t\geqslant 0}$，如果其状态空间为 N，它的转移概率函数 $P_{ij}(t)$ 满足(28.3)，则称为**一维连续时间马尔可夫分支过程**(one dimentional continuous time Markov branching process).

28.2.1 母函数

设 $\{Z_t\}_{t\geqslant 0}$ 为一维连续马尔可夫分支过程，$\{P_k\}_{k\in N}$ 是个体的后代分布，a 为寿命分布的参数，$P_{ij}(t)$ 为转移函数，记：

$$f(s)=\sum_{j=0}^{\infty}P_j s^j,|s|\leqslant 1,$$
$$u(s)=a[f(s)-s],$$
$$F(s,t)=\sum_{k=0}^{\infty}P_{1k}(t)s^k,|s|\leqslant 1.$$

本节仍假设 $Z_0=1$，且不考虑平凡情况.
此时成立下面的母函数方程：

$$F(s,u+t)=F(F(s,t),u).$$

定理 28.2.2 若对充分小的 $\varepsilon>0$，

$$\left|\int_{1-\varepsilon}^{1}\frac{\mathrm{d}s}{u(s)}\right|=\infty,$$

则对 $\forall\ t>0,F(1,t)=1$；否则 $F(1,t)<1,\forall\ t>0$.

28.2.2 灭绝概率和矩

令 $q=P(Z_t\to 0,t\to\infty)$ 表示灭绝概率，则有

定理 28.2.3 q 是方程 $u(s)=0$ 在$[0,1]$内的最小根. Z_t 的前二阶矩为

$$EZ_t=\mathrm{e}^{\lambda t},\lambda=u'(1).$$
$$EZ_t^2=\begin{cases}u''(1)a^{-1}(\mathrm{e}^{2\lambda t}-\mathrm{e}^{\lambda t}), & a\neq 0,\\ u''(1)t, & a=0.\end{cases}$$

28.2.3 嵌入 GW 过程

设 $\{Z_t\}_{t\geqslant 0}$ 为一维连续时间马尔可夫分支过程，母函数为 $F(s,t)$，那么有

定理 28.2.4 $\forall\ \delta>0$，序列 $\{Z'_n\}_{n\in N}=\{Z_{n\delta}\}_{n\in N}$ 是一个 GW 过程，它的母函数 $f(s)=F(s,\delta)$.

于是对 $\{Z_t\}$ 的许多性质可以转化为 GW 过程 $\{Z'_n\}$ 的描述，后

者是一个嵌入过程(由连续时间过程得到的离散过程,叫原过程的嵌入).

28.2.4 极限性质

$\{Z_t\}$的极限性质可以由嵌入 GW 过程得到,这里叙述一些于后,注意到 $\lambda=a(m-1)$,因此过程的分类可以用 λ 改述.

超临界情形

定义过程$\{W_t\}_{t\geqslant 0}=\{Z_t e^{-\lambda t}\}_{t\geqslant 0}$,则它关于自然 σ 代数族(即由 W 产生的 σ 代数族)是非负鞅,以 W 表示 $t\to\infty$时的 a.s. 极限.(此结论对一切 λ 值亦对),则有:

定理 28.2.5 若 $\lambda>0$,那么 $P(W=0)=1$ 或 $EW=1$,后者成立$\Leftrightarrow \sum\limits_{k=2}^{\infty}KP_k\log K<\infty$.

此外当前者成立时有:

(1) $P(W=0)=q$.

(2) 存在$(0,\infty)$上的连续密度函数 $W(x)$,使 $\forall$ $0<x_1<x_2<\infty$,

$$P(x_1\leqslant W\leqslant x_2)=\int_{x_1}^{x_2}W(x)\mathrm{d}x.$$

临界情形

定理 28.2.6 若 $\lambda=0,\sigma^2=a\left[\sum\limits_{k=0}^{\infty}K^2P_k-1\right]<\infty$,则

$$\lim_{t\to\infty}P\left\{Z_t>\frac{\sigma^2}{2}tx\,|\,Z_t>0\right\}=e^{-x}.$$

次临界情形

定理 28.2.7 若 $\lambda<0$,则

$$\lim_{t\to\infty}P\{Z_t=j\,|\,Z_t>0\}=b_j,j\in N,$$

存在,且 $b_j\geqslant 0,\sum\limits_{j=0}^{\infty}b_j=1$.

28.3 依龄分支过程

在(28.2)的过程中,如果个体的寿命分布为一般分布函数 $G(t)$,那么相应的过程 $\{Z_t\}_{t\geqslant 0}$ 一般不是马尔可夫过程,但它仍具有一些类似于马尔可夫分支过程的性质. 由于 $G(t)$ 与年龄有关,此过程称为**依龄分支过程**.

28.3.1 母函数基本方程

对 $t\geqslant 0$,记 $P_k(t)=P(Z_t=k)$, $k\in N$,考虑母函数 $F(s,t)=\sum_{k=0}^{\infty}P_k(t)s^k$, $|s|\leqslant 1$.

定义 28.3.1 母函数基本方程 设 $\{P_k\}$ 和 $f(s)$ 同(28.2),则

$$F(s,t)=s[1-G(s)]+\int_0^t f(F(s,t-y))\mathrm{d}G(y),\ |s|\leqslant 1.$$

定理 28.3.2 设 f 是一概率母函数,G 为 $[0,+\infty)$ 上的分布函数且 $G(0_+)\triangleq\lim\limits_{t\downarrow 0}G(t)=0$,则(28.3.1)有一个解 $F(s,t)$,它对固定的 t 是一母函数,且为唯一的有界解.

28.3.2 灭绝概率和矩

$F(s,t)$ 如定义 28.3.1,其余记号沿用(28.2),我们有($Z_0=1$):

定理 28.3.3 (1) 若 $0\leqslant s\leqslant q$,则 $\uparrow\lim\limits_{t\to\infty}F(s,t)=q$;若 $q\leqslant s<1$,则 $\downarrow\lim\limits_{t\to\infty}F(s,t)=q$,这里 $\uparrow(\downarrow)$ 表单增(减)极限.

(2) $P(Z_t\neq 0,\forall\, t\geqslant 0)>0\Leftrightarrow m>1$,若后一条件成立,则概率 $P(Z_t=0$,对某 $t>0)$ 是方程 $f(s)=s$ 的唯一小于 1 的非负根.

关于依龄分支过程和链(branching process (chian) dependent on ages),有

定理 28.3.4 (1) $EZ_t = \mu_t$ 是下列方程的唯一解：

$$\mu(t) = [1 - G(t)] + m\int_0^t \mu(t-y)\mathrm{d}G(y).$$

它在有限区间上是有界的.

(2) $EZ_{t_1}Z_{t_2} = \mu_2(t_1, t_2)$是下列方程的唯一解(有限区间上有界)：

$$\mu_2(t_1, t_2) = r(t_1, t_2) + m\int_0^{t_1} \mu_2(t_1 - y, t_2 - y)\mathrm{d}G(y),$$

其中

$$r(t_1, t_2) = 1 - G(t_2) + m_2\int_0^{t_1} \mu(t_1 - y)\mu(t_2 - y)\mathrm{d}G(y)$$

$$+ m\int_{t_1}^{t_2} \mu(t_2 - y)\mathrm{d}G(y).$$

而 $$m_2 = E(Z_1(Z_1 - 1)).$$

28.4 多维离散分支过程

考虑群体有 m 种不同类型的个体，分别以 $1, 2, \cdots, m$ 表示其型号，每隔一单位时间，型号为 i 的个体以概率 $P^{(i)}(j_1, \cdots, j_m)$分裂成为 j_1 个第 1 型个体，$\cdots$，j_m 个第 m 型个体，此分裂对不同型号或同型号不同个体之间都是独立的，以 Z_{n_i}表时刻 n(或第 n 代)群体所含的第 i 型个体总数，$Z_n = \{Z_{n_1}, \cdots, Z_{n_m}\}$，则$\{Z_n\}_{n\in N}$是一个 m 维随机序列. 多维分支过程很好地描述了这种演化.

28.4.1 若干定义

设 $\boldsymbol{s} = (s_1, \cdots, s_m) \in [0, 1]^m$，$\boldsymbol{j} = (j_1, \cdots, j_m) \in N^n$，$\boldsymbol{0} = (0, \cdots, 0)$，$\boldsymbol{1} = (1, \cdots, 1)$，为 m 维向量；对 $i = 1, \cdots, m$，$P^{(i)}(\boldsymbol{j}) = P^i(j_1, \cdots, j_m)$是 N^m 上的概率分布，其母函数为 $f^{(i)}(\boldsymbol{s}) = f^{(i)}(s_1, \cdots, s_m) =$

$$\sum_{j_1,\cdots,j_m\geqslant 0} P^{(i)}(j_1,\cdots,j_m)s_1^{j_1}\cdots s_m^{j_m} = \sum_{j\in N^m} P^{(i)}(j)s^j,$$

记
$$\boldsymbol{P}(\boldsymbol{j}) = (P^{(1)}(\boldsymbol{j}),\cdots,P^{(m)}(\boldsymbol{j})),\boldsymbol{f}(\boldsymbol{s}) = (f^{(1)}(\boldsymbol{s}),\cdots,f^{(m)}(\boldsymbol{s})),$$

则
$$\boldsymbol{f}(\boldsymbol{s}) = \sum_{j\in N^m} \boldsymbol{P}(\boldsymbol{j})\boldsymbol{s}^{\boldsymbol{j}}, \tag{28.4}$$

定义 28.4.1 状态空间为 N^m 的马尔可夫链,若其转移函数 $P(\boldsymbol{i},\boldsymbol{j})=P(Z_{n+1}=\boldsymbol{j}|Z_n=\boldsymbol{i}),\boldsymbol{i},\boldsymbol{j}\in \boldsymbol{N}^m$ 为$[\boldsymbol{f}(\boldsymbol{s})]^{\boldsymbol{i}}$ 中 $\boldsymbol{s}^{\boldsymbol{j}}$ 的系数,这里 $\boldsymbol{f}(\boldsymbol{s})$ 具有(28.4)的形式. $\{Z_n\}_{n\in N}$ 称为 m 维离散分支过程(m-dimentional discrete time branching process),也称为 GW 过程.

$\boldsymbol{f}(\boldsymbol{s})$将起一维时 $f(s)$的作用.

注意,若存在一元素非负的 m 维矩阵 $\boldsymbol{M}$ 使 $\boldsymbol{f}(\boldsymbol{s})=\boldsymbol{M}\boldsymbol{s}^T$,则此时的分支过程变为普通马尔可夫链,称此过程是奇异的(singular),下面假设考虑的分支过程是非奇异的.

28.4.2 矩

记 $e_i=(\delta_{ij},j=1,\cdots,m),i=1,\cdots,m,m_{ij}=E(Z_{1j}|Z_0=e_i)$, $\boldsymbol{M}=(m_{ij})_{i,j\in N}$,假定 $\boldsymbol{M}$ 中的元都是有限的,它们可由 $\boldsymbol{f}(\boldsymbol{s})$求导得到,一般的 Z_n 的矩可由 $\boldsymbol{M}$ 表示:$E(Z_n|Z_0)=Z_0\boldsymbol{M}^n$.

定义 28.4.2 m 维 GW 过程$\{Z_n\}_{n\in N}$如果相应的均值矩阵 $\boldsymbol{M}$ 是严格正的,即存在 $n\geqslant 1$,使 $\boldsymbol{M}^n$ 的元素严格正,称它是**正规则的**(positive regular).

由矩阵论知对严格正矩阵,存在一个最大的正特征根,它为一单根.

28.4.3 灭绝概率和非常返性

设$\{Z_n\}_{n\in N}$是 m 维 GW 过程,$\boldsymbol{f}(\boldsymbol{s})$为它的母函数. 状态 $\boldsymbol{0}$ 是吸

引态,对其余的状态有

定理 28.4.3 若$\{Z_n\}_{n\in N}$是正规则和非奇异的,那么对 $j\in N^m/\{0\}$,有

$$p(Z_n = j,\text{无穷多个 } n) = 0$$

即所有非零状态都是暂留的.

考虑灭绝概率,记

$$q_i = p(Z_n = 0,\text{对某 } n \in N | Z_0 = e_i),$$
$$\boldsymbol{q} = (q_1,\cdots,q_m).$$

定理 28.4.4 设$\{Z_n\}_{n\in N}$是正规则和非奇异的,ρ 为 M 的最大特征根.则有

(1) 若 $\rho\leqslant 1$,则 $\boldsymbol{q}=\mathbf{1}$;若 $\rho>1$,则 $\boldsymbol{q}<\mathbf{1}$.

(2) $\boldsymbol{q}$ 是方程组 $\boldsymbol{f}(\boldsymbol{s})=\boldsymbol{s}$ 在$[0,1]^m$ 内的唯一解.

28.5 多维连续时间分支过程

在(28.4)所考虑的群体中,若每个个体生存一随机时间后消失,同时产生的后代个数也是随机的,那么描述此现象的模型是一个时间连续的随机过程.和一维情形相类似,个体的寿命分布是否为指数型(参数与个体的类型有关),决定过程的马氏性.这里考虑马尔可夫分支过程.

定义 28.5.1 定义在概率空间$\langle\Omega,\mathscr{F},P\rangle$上取值于 N^m 的强马尔可夫过程$\{Z_t\}_{t\geqslant 0}$,如果它的转移函数 $P(\boldsymbol{i},\boldsymbol{j};t)$满足**分支条件**(branch condition)

$$\sum_{\boldsymbol{j}\in N^m} P(\boldsymbol{i},\boldsymbol{j};t)\boldsymbol{s}^{\boldsymbol{j}} = \prod_{k=1}^{m}\Big[\sum_{\boldsymbol{j}\in N^m} P(e_k,\boldsymbol{j};t)\boldsymbol{s}^{\boldsymbol{j}}\Big]^{i_k};$$

其中 $\boldsymbol{i} = (i_1,\cdots,i_m) \in N^m,\boldsymbol{s} \in [0,1]^m,t \geqslant 0.$

称它为 m **维连续时间分支过程**(m-dimensional continuous time branching process).

和一维连续时间情形类似，设第 i 型每个个体的寿命分布服从参数为 a_i 的指数分布，产生后代的概率分布为 $\{p^{(i)}(j_1,\cdots,j_m)\}$，则由 $\boldsymbol{a}=(a_1,\cdots,a_m)$，$\boldsymbol{p}(\boldsymbol{j})=(p^{(1)}(\boldsymbol{j}),\cdots,p^{(m)}(\boldsymbol{j}))$ 求出密度矩阵 $\boldsymbol{Q}$，从而构造出最小 $\boldsymbol{Q}$ 过程. 也可由解母函数的向前和向后方程求得分支转移函数和母函数，详细的可参阅[40].

28.6 一般马尔可夫分支过程

考虑一个系统，它由若干粒子构成，每个粒子独立地按照某种马尔可夫过程运动，而在过程的一定状态下，它消失或分裂的后代个数是随机的，这些下一代粒子遵循先辈的同样运动，如此下去. 在此种场合不但关心粒子的总数和年龄，而且需要考虑粒子在相空间的位置. 这种考虑导致一般马尔可夫分支过程的研究.

现叙述一般马尔可夫分支过程定义.

设 $(X,\mathscr{B})$ 是一可测空间，称为过程的**型空间**(type space)，$\theta\notin X$ 为一附加点，记 $X^{(0)}=\{\theta\}$，对 $n\geqslant 1$，$X^{(n)}$ 表示 X 的 n 次幂 X^n 的对称化，即 $X^{(n)}$ 的一个点对应于 X^n 中这样一个集合，此集合中点的坐标仅顺序不同. $X^{(n)}$ 中点将用 $\langle x_1,\cdots,x_n\rangle$ 表示，其中 $x_i\in X$. 令 $\hat{X}=\bigcup\limits_{n=0}^{\infty}X^{(n)}$，以 $\hat{\mathscr{B}}$ 表示 $\hat{X}$ 上由 $\mathscr{B}$ 引入的 σ 代数. 对 $A\subset X$，以 $1_A(\cdot)$ 表 A 的示性函数，设 $\hat{x}\in\hat{X}$，记

$$\hat{x}(A)=\begin{cases}0, & \hat{x}=\theta,\\ \sum\limits_{k=1}^{n}1_A(x_k), & \hat{x}=\langle x_1,\cdots,x_n\rangle,\end{cases}$$

它代表 $\hat{x}$ 位于 A 中的粒子数($\hat{x}$ 视为某时刻系统的状态).

定义 28.6.1 一个 $\langle\hat{X},\hat{\mathscr{B}}\rangle$-值马尔可夫过程，如果它的转移函数 $P(t,\hat{x};\hat{A})$ 具有以下性质：

(1) $P(t,\theta;X^{(0)})=1,t\geqslant 0$.

(2) 对任意 $t \geqslant 0$ 以及 X 的任一可测分解：$X = A_1 + A_2 + \cdots + A_m$，关于所有非负整数 $n_j, j = 1, \cdots, m, \hat{x} = \langle x_1, \cdots, x_k \rangle \in \hat{X}, k > 0$，有

$$P(t, \langle x_1, \cdots, x_k \rangle; \{\hat{x}[A_j] = n_j, j = 1, \cdots, m\})$$
$$= \sum_{\substack{n_{j_1} + \cdots + n_{j_m} \\ j = 1, \cdots, m}} \prod_{\nu=1}^{k} P(t, \langle x_\nu \rangle; \{\hat{x}[A_j] = n_{j\nu}, j = 1, \cdots, m\})$$

称它为马尔可夫分支过程(Markov branching process).

在上述定义中,(1)表示没有粒子自系统外流入,而(2)中的等式表示独立分支性.

引入 $\hat{X}$ 的一个直观是：以 Z_t 表示时间 t 系统的粒子数，X_t^i 表示第 i 个粒子于 t 时的位置，则 t 时系统有状态 $X_t = (X_t^1, \cdots, X_t^{Z_t})$，因为考虑同型粒子是不可分辨的，故取 $\hat{X}$ 为系统的状态空间.

29 鞅论

29.1 定义与例

鞅论创立于20世纪30年代末至50年代初杜布(J L Doob)和列维(P Lévy)的工作,形成随机过程论的一个独立分支.60年代初迈耶(P A Meyer)解决了上鞅分解问题,并发展了平方可积鞅理论.库尼塔(H Kunita)和瓦塔奈伯(S Watanabe)对平方可积鞅的研究、迈耶和戴拉谢利(C Dellacherie)创立的随机过程一般理论,使现代鞅论有了开创性的发展,成为70年代以来随机过程论中最活跃和最富成果的分支之一.鞅与随机积分理论被愈来愈广泛地应用于马尔可夫过程、点过程、估计理论和随机控制等理论分支及应用领域.

设T为任意的有序整数集或实数集,分别对应于离散时间与连续时间.$(\mathscr{F}_t)_{t\in T}$为$\mathscr{F}$的一个非降子σ代数族,$X_T\triangleq\{X_t,t\in T\}$为基本概率空间$(\Omega,\mathscr{F},P)$上的随机过程.

定义 29.1.1 设$(\mathscr{F}_t)$适应过程X_T满足

(1) $E|X_t|<\infty$,

(2) $E(X_t|\mathscr{F}_s)=X_s$, a.s., $s<t$, $s,t\in T$,

则称过程X_T为$(\mathscr{F}_t)$**鞅**(martingale),其中$(\mathscr{F}_t)$为参考族.如若X_T满足

(3) $EX_t^+<\infty$, $[EX_t^-<\infty]$,

(4) $E(X_t|\mathscr{F}_s)\leqslant[\geqslant]X_s$, a.s., $s<t$, $s,t\in T$,

则称过程X_T为$(\mathscr{F}_t)$**上鞅**(supermartingale)[**下鞅**(submartingale)].

今后,方括号内陈述者表示替代前方概念或命题而可平行建立的相应的概念或命题.

当 T 为实数轴上区间时,以下恒设鞅、上鞅和半鞅,是可分的.

由定义知 $\mathscr{F}_t \supset \mathscr{N}_t \triangleq \sigma\{X_s, s \leqslant t\}$,有时就取 $\mathscr{N}_t$ 代替定义中的 $\mathscr{F}_t$,相应的条件期望写为 $E(X_t | \mathscr{N}_s)$ 或 $E(X_t | X_\tau, \tau \leqslant s)$,离散时有 $E(X_{n+k} | \mathscr{N}_n)$ 或 $E(X_{n+k} | X_1, X_2, \cdots, X_n)$. 恒设 $s, t \in T$.

由条件数学期望定义知定义 29.1.1/(2)及 29.1.1/(4)分别有如下等价形式

(1) $\int_A X_t \mathrm{d}P = \int_A X_s \mathrm{d}P, \quad \forall A \in \mathscr{F}_s,$

(2) $\int_A X_t \mathrm{d}P \geqslant [\leqslant] \int_A X_s \mathrm{d}P, \quad \forall A \in \mathscr{F}_s.$

例 29.1.2 设 $\xi_n, n \geqslant 1$ 为独立 r.v. 列, $E\xi_n = 0$(期望定义中含 $E|\xi_n| < \infty$). 令 $\mathscr{F}_n = \sigma\{\xi_i, i \leqslant n\}$,即由 $\xi_1, \xi_2, \cdots, \xi_n$ 生成的 σ 代数,又令 $X_n = \sum_1^n \xi_i$,则 $(X_n)_{n \geqslant 1} \triangleq \{X_n, n \geqslant 1\}$ 为 $(\mathscr{F}_n)$ 鞅.

事实上,由条件期望性质、$\{\xi_n\}$独立性及 $E\xi_n = 0$, $E(X_{n+1} | X_1, \cdots, X_n) = E(X_n | X_1, \cdots, X_n) + E(\xi_{n+1} | X_1, \cdots, X_n) = X_n + E(\xi_{n+1} | \xi_1, \cdots, \xi_n) = X_n + E\xi_{n+1} = X_n$, a.s.

进一步尚可证 $E\xi_n = 0$ 是充要条件,只要关于 ξ_i 的其他条件不变.

例 29.1.3 设 $(\xi_n)_{n \geqslant 1}$ 是 r.v. 列(未必独立), $E|\xi_n| < \infty$ 且 $E(\xi_{n+1} | \xi_1, \cdots, \xi_n) = 0, n > 1$,则 $(X_n)_{n \geqslant 1}$ 为 $(\mathscr{F}_n)$鞅,这里 $X_n, \mathscr{F}_n$ 同例 29.1.2,此鞅也称杜布(Doob)过程.

例 29.1.4 设 r.v. ξ 有 $E|\xi| < \infty$, $(\mathscr{F}_t)$为参考族. 令 $X_t = E(\xi | \mathscr{F}_t), t \in T$,则 X_T 为$(\mathscr{F}_t)$鞅.

例 29.1.5 设$(X_n)_{n \geqslant 1}$为下鞅,则$(X_n^+)_{n \geqslant 1}$也是,这里 $X_n^+ = X_n \vee 0$. 又若 $X_n \geqslant 0, E|X_n|^r < \infty$,则$(|X_n|^r)_{n \geqslant 1}$也是下鞅, $r \geqslant 1$.

证 由 $X_n^+ = \frac{1}{2}(X_n + |X_n|)$知为 $\mathscr{F}_n \triangleq \sigma\{X_n\}$可测,且为升函

数. 故由定义 29.1.1/(4), 令 $g(x)=x^+$, 有 $g(X_n)\leqslant g(E(X_{n+1}|X_1,\cdots,X_n))\leqslant E(g(X_{n+1}|X_1,\cdots,X_n))$. 此即 $X_n^+\leqslant E(X_{n+1}^+|X_1,\cdots,X_n)$. 由 $\sigma\{X_i^+,i\leqslant n\}\subset\sigma\{X_i,i\leqslant n\}$ 及有关性质知 $X_n^+\leqslant E(X_{n+1}^+|X_1^+,\cdots,X_n^+)$, 从而 $(X_n^+)_{n\geqslant1}$ 是下鞅. 后半结论可仿证.

例 29.1.6 设 $\xi_n,n\geqslant1$ 为独立列, $E\xi_n=1,n\geqslant1$. 令 $Y_n=\prod_{i=1}^{n}\xi_i$, 则 (Y_n) 为 $(\mathscr{F}_n)$ 鞅, 这里 $\mathscr{F}_n$ 定义同例 29.1.2. 事实上, 令 $X_n^\pm\triangleq\ln Y_n^\pm$, 则由例 29.1.2 及 $x=\ln y$ 为 $y(>0)$ 的升函数, 依鞅的定义, 可证 (Y_n) 为 $(\mathscr{F}_n)$ 鞅.

例 29.1.7 直线上伯努利游动中的鞅 设 $(\xi_n)_{n\geqslant1}$ 是独立伯努利 r. v. 列, $P(\xi_n=1)=p$, $P(\xi_n=-1)=q\triangleq1-p$. 令 $S_n=\sum_{1}^{n}\xi_i$, 它是 n 次游动后质点的位移. 则 $X_n\triangleq(q/p)^{S_n},n\geqslant1$ 是 $(\mathscr{F}_n)$ 鞅, 这里 $\mathscr{F}_n=\sigma\{\xi_i,i\leqslant n\}$. 此结论利用例 29.1.6, 注意 $E(q/p)^{\xi_i}=1$ 可立即证得. 此外, 尚可知 $EX_n=1$.

例 29.1.8 与布朗运动的伊藤积分有关的鞅 设 (B_t) 是布朗运动, 对 (B_t) 伊藤随机积分过程 $M_t=\int_0^t f(s)\mathrm{d}B_s,0\leqslant t\leqslant T$, 在当 f 为初等函数, 或更一般地 f 为 $(\mathscr{F}_t)$ 可测, 且 $\int_0^T E|f(s)|^2\mathrm{d}s<\infty$ 时, (M_t) 是鞅.

例 29.1.9 可加泛函中的例子 设 (A_t) 为可加泛函过程, 且满足 $E_x|A_t|<\infty,E_xA_t=0,t\geqslant0,\forall\ x\in R$. 则 (A_t) 是 $(\mathscr{F}_t)$ 鞅. 由此若 (X_t) 为扩散过程, A 为无穷小算子, $u\in D_A$, 则下列过程 (Y_t) 为鞅 [70], $Y_t=u(X_t)-u(X_0)-\int_0^t Au(X_s)\mathrm{d}s,t>0$.

注 29.1.10 鞅的实际背景 设某企业第 n 年产值 X_n 元, $E(X_{n+1}|X_1,\cdots,X_n)$ 是已知前 n 年产值情况下预计第 $n+1$ 年的平均产值. 如以 X_n 作第 $n+1$ 年的本金投入, 则当 $E(X_{n+1}|X_1,\cdots,$

$X_n)=X_n$ 时，表示无盈亏，而 F 鞅则表示有盈利.

例 29.1.3 中 ξ_n 如表示对策(game)之第 n 局实现的增益(可 $\leqslant 0$)，则 $E(\xi_{n+1}|\xi_1,\cdots,\xi_n)=0$ 表示对策是无损的. 如 τ 为($\mathscr{F}_n$)停时(stopping time)，则由公式 29.7.2 知 $E\left(\sum_{1}^{\tau}\xi_i\right)=0$，这表明对停时的选择，不能增加无损系统的平均增益.

29.2 鞅的基本性质

性质 29.2.1 初等性质

(1) 设 X_T 为鞅[上鞅、下鞅]，r.v. $\eta>0$ 且为 $\mathscr{F}_s$ 可测，则对 $\forall\ t>s$，有

$$E(X_t\eta)=[\geqslant、\leqslant]E(X_s\eta).$$

(2) 如 X_T 为鞅，则 $EX_t=\text{const}$，且 $E|X_t|\uparrow$；如 X_T 为下鞅，则 $EX_t\uparrow$.

性质 29.2.2 判别性质

(1) 如 X_T 为下[上]鞅，则 $-X_T\triangleq\{-X_t,t\in T\}$ 为上[下]鞅.

(2) 如 X_T 为下[上]鞅，f 为连续不减凸(凹)函数，$Ef(X_t)<\infty,t\in T$，则 $f(X_T)\triangleq\{f(X_t),t\in T\}$ 是下[上]鞅；如 X_T 为鞅，f 为连续凸函数，$Ef(X_t)<\infty,t\in T$，则 $f(X_T)$ 是下鞅，但未必是鞅.

(3) 如 X_T 是下鞅，则 $\{(X_t-c)^+,t\in T\}$ 也是.

(4) 如 X_T,Y_T 为两个鞅[上鞅]，则 $(X+Y)_T$ 是鞅[上鞅]，而 $(X\wedge Y)_T$ 是上鞅.

公式 29.2.3 几个不等式

(1) **下鞅极大不等式**(maximum inequatity for submartingale) 设 X_T 是下鞅，则 $P(\sup\limits_{t\in T}X_t^+\geqslant c)\leqslant\frac{1}{c}\sup\limits_{t\in T}EX_t^+$，$c$ 为任一常数.

(2) 科莫戈罗夫不等式　设 X_T 是鞅, $\sup E|X_t|^p<\infty$, $p\geqslant 1$, 则

$$P(\sup_{t\in T}|X_t|\geqslant c)\leqslant\frac{1}{c^p}\sup_{t\in T}E|X_t|^p, c>0.$$

又若 T 有最大元 t^*, 则上两不等式右方在 t^* 达到上确界.

(3) 杜布不等式　设 X_T 是下鞅, 其几乎所有轨道为右连续, 则

$$E(\sup X_t^+)\leqslant\frac{e}{e-1}(1+\sup E(X_t^+\log^+X_t^+)),$$

$$E((\sup X_t^+)^p)\leqslant\left(\frac{p}{1-p}\right)^p\sup E(X_t^+)^p,\quad p>1.$$

(4) 杜布上穿不等式(upcrossing inequatity)　设 $\nu[a,b)$ 为下鞅 X_T 从下向上穿过区间 $[a,b]$ 的次数, 则

$$E\nu[a,b)\leqslant\sup E(X_t-a)^+/(b-a).$$

定理 29.2.4　任意抽样定理与杜布停止定理

(1) 离散时的任意抽样定理(optional sampling theorem) 设 (X_n) 是鞅[上鞅], τ 为停时, 如 $P(\tau<\infty)=1$, 且 $E(\sup_{n\geqslant 0}|X_{\tau\wedge n}|)<\infty$, [$\exists$ r.v. $W\geqslant 0$, 满足 $EW<\infty$ 及对 $\forall\ n, X_{\tau\wedge n}>-W$], 则

$$EX_\tau=[\leqslant]EX_0.$$

(2) 设 (X_n) 对 $\sigma\{Y_0,\cdots,Y_n\}$, $n\geqslant 0$ 为鞅, τ 为 (Y_n) 停时, 如 $E\tau<\infty$ 且 $\exists$ 常数 $k<\infty$ 使 $E(|X_{n+1}-X_n||Y_0,\cdots,Y_n)\leqslant k$, $\forall\ n<\tau$, 则

$$EX_\tau=EX_0.$$

(3) 任意停止定理(optional stopping theorem)　设 (X_n) 为鞅, τ 为停时, 且

1) $P(\tau<\infty)=1$,　2) $E|X_\tau|<\infty$,　3) $\lim_{n\to\infty}E(X_nI_{(\tau>n)})=0$,

则

$$EX_\tau=EX_0.$$

(4) 连续时的任意抽样定理　设 (X_t) 为 $(\mathscr{F}_t)$ 鞅[下鞅], τ 为 $(\mathscr{F}_t)$ 停时, 如 $P(\tau<\infty)=1$, 且 $E(\sup|X_t|)<\infty$, [$(X_{t\wedge\tau})^-$, $t\geqslant 0$

为一致可积],则

$$EX_\tau=[\geqslant]EX_0.$$

(5) 杜布停止定理　设 X_{R^+} 为($\mathscr{F}_t$)鞅[上鞅],τ_i 为($\mathscr{F}_t$)停时,$i=1,2$,($\mathscr{F}_t$)右连续

1) 如 $\tau_1\leqslant\tau_2$,有 $E(X_{\tau_2}|\mathscr{F}_{\tau_1})=[\leqslant]X_{\tau_1}$,a.s.;

2) (一般情形)则 $E(X_{\tau_2}|\mathscr{F}_{\tau_1})=[\leqslant]X_{\tau_1\wedge\tau_2}$,a.s..

29.3　鞅的闭合与收敛性

定义 29.3.1　如 $\{X_t,t\in T,Y\}$ 是($\mathscr{F}_t$)鞅[下鞅],换言之,如 $E|Y|<\infty$,$Y\in\mathscr{F}\triangleq\sigma\{\mathscr{F}_t,t\in T\}$ 且 $E(Y|\mathscr{F}_t)=[\geqslant]X_t$,a.s.$\forall t\in T$,称($\mathscr{F}_t$)鞅[下鞅]$X_T$ 对($\mathscr{F}_t$)**右闭于** Y(rightly closed by Y),此 Y 称为**右闭元**.

如 T 有最大元 t^*,则 X_{t^*} 为右闭元.

定义 29.3.2　如对每一 $t\in T$,X_t 可积,且 $c\to+\infty$ 时 $\int_{(X_t>c)}|X_t|\mathrm{d}P$ 对 $t\in T$ 一致地趋于零,称过程(X_t)为**一致可积**的.

性质 29.3.3　一致可积的等价条件　设 $E|X_t|^p<\infty$,则(X_t)一致可积$\Leftrightarrow\sup\limits_{t\in T}E|X_t|^p<+\infty$　($p\geqslant1$).当 $p=1$ 时尚要求均匀可积,即对$\forall\ \varepsilon>0$,$\exists\ \delta>0$ 使当 $P(A)\leqslant\delta$,$A\in\mathscr{F}$ 时,$\int_A|X_t|\mathrm{d}p\leqslant\varepsilon$,$\forall\ t\in T$.

定理 29.3.4　设 X_T 为($\mathscr{F}_t$)鞅或下鞅.下述等价:

(1) X_T 为($\mathscr{F}_t$)右闭,

(2) X_T 一致可积,

(3) 在 L_1 收敛意义下存在 $\lim\limits_{t\uparrow t^*}E|X_t-Y|=0$.

如三者中有一成立,则 $\lim\limits_{t\uparrow t^*}X_t=Y$ 存在(a.s.),且 Y 即 X_T 的右闭元.

定理 29.3.5 下鞅的右正则化 设区间 $T\subset\overline{R_+}$，X_T 为 $(\mathscr{F}_t)$ 下鞅，则 $\{X_{t+0},\mathscr{F}_{t+0},t\in T\}$ 也是，且样本右连续[P]. 此时在使 EX_t 连续且 $\mathscr{F}_{t+0}=\mathscr{F}_t$ 的每一点 t 上，$X_t=X_{t+0}$，$X_{t^*+0}=X_{t^*}$ [P].

取 $T=R_+\triangleq[0,\infty)$，$\overline{R_+}=[0,\infty]$. 关于鞅的收敛性，有

定理 29.3.6 设 X_{R_+} 为 $(\mathscr{F}_t)$ 上鞅，几乎所有轨道右连续. 如 $\sup EX_t^-<\infty$（等价地 $\sup E|X_t|<\infty$）则∃ 某可积 r.v. X_∞，使 $X_t\xrightarrow{\text{a.s.}}X_\infty$，$t\to\infty$. 又若 X_{R_+} 为非负 $(\mathscr{F}_t)$ 上鞅，则 $X_{\overline{R_+}}$ 也是.

定理 29.3.7 设 X_{R_+} 为一致可积鞅[上鞅]，且几乎所有轨道右连续，则当 $t\to\infty$ 时有 $X_t\xrightarrow{\text{a.s. 且 } L_1}X_\infty$，且 $X_{\overline{R_+}}$ 为鞅[上鞅].

由该定理可推得如下定理.

定理 29.3.8 设 X_{R_+} 为鞅[非负下鞅]，轨道右连续(a.s.)，对某 $p>1$，$\sup E|X_t|^p<\infty$，则 X_{R_+} 为一致可积，且 $X_t\xrightarrow{\text{a.s.},L^p}X_\infty$，$\|X_\infty\|_p=\sup\|X_t\|_p$.

定理 29.3.9 设 $(X_t)_{t>0}$ 为 $(\mathscr{F}_t)_{t>0}$ 上鞅，轨道右连续(a.s.). 如 $\sup\limits_{t>0}EX_t<\infty$，则 $t\downarrow 0$ 时 $X_t\xrightarrow{\text{a.s.},L_1}X_0$，$X_0$ 为某 r.v.，且对 $\mathscr{F}_0\triangleq\bigcap\limits_{t>0}\mathscr{F}_t$ 可测、可积. 此外 X_{R_+} 为 $(\mathscr{F}_t)$ 上鞅.

29.4 上鞅与下鞅的分解

29.4.1 里斯分解

定义 29.4.1 如 $\lim\limits_{T\ni t\to\infty}E\pi_t=0$，称 $(\mathscr{F}_t)$ 非负上鞅 π_T 为**位势**(potential)，当 T 连续时还要求 X_T 右连续[P].

定义 29.4.2 设 X_T 为上鞅，如存在鞅 M_T 及位势 π_T，使 $X_t=M_t+\pi_t$，$t\in T$，则称 X_T 有**里斯分解**(Riesz decomposition)，当 T 连续时还要求 X_T 几乎所有轨道右连续.

定理 29.4.3 设(X_t)为$(\mathscr{F}_t)$上鞅，当T连续时，几乎一切轨道右连续且$(\mathscr{F}_t)$右连续. 则

(1) X_T如有里斯分解，则在随机等价意义下必唯一.

(2) X_T有里斯分解$\Leftrightarrow \lim\limits_{t\to\infty} EX_t > -\infty$.

(3) X_T非负，$X_T = M_T + \pi_T$为其里斯分解，则M_T为非负鞅.

(4) 如(X_t)一致可积，必存在$\lim\limits_{t\to\infty} X_t \overset{L_1}{=} X_\infty$，且$E(X_\infty | \mathscr{F}_t)$，$t \in \overline{R_+}$为鞅. 作右连续修正后，$X_{\overline{R_+}}$有里斯分解：

$$X_t = E(X_\infty | \mathscr{F}_t) + (X_t - E(X_\infty | \mathscr{F}_t)).$$

29.4.2 杜布分解

定义 29.4.4 如$\alpha_0 = 0$，$\alpha_s \leqslant \alpha_t [P]$，一切$s \leqslant t$；则称右连续$(\mathscr{F}_t)$过程$\alpha_{R_+}$为$(\mathscr{F}_t)$**递增的**(increasing). 如对一切以概率1有界、非负且有左极限的$(\mathscr{F}_t)$鞅X_{R_+}有

$$E\int_0^\infty X_{t-0} \mathrm{d}\alpha_t = EX_\infty \alpha_\infty \triangleq E(\lim_{t\to\infty} X_t \cdot \lim_{t\to\infty} \alpha_t).$$

则称之为自然的(natural).

"自然"的概念，出现早于"可料"(定义 29.6.5). 由于自然增过程就是可料增过程，"自然"这一概念已为"可料"概念代替.

定义 29.4.5 如$\sup E\alpha_t < \infty$，称$(\mathscr{F}_t)$增过程α_{R_+}为一致可积的.

定理 29.4.6 一致可积增过程α_{R_+}是自然的，当且仅当对$\forall t > 0$及任意右连续且有左极限的有界鞅X_{R_+}，有

$$E\int_0^t X_s \mathrm{d}\alpha_s = E\int_0^t X_{s-0} \mathrm{d}\alpha_s.$$

定理 29.4.7 杜布-迈耶(Doob-Meyer)分解 设π_{R_+}为$(\mathscr{F}_t)$右连续势，使$\{\pi_\tau, \tau \in \mathscr{T}\}$一致可积，其中$\mathscr{T} \triangleq \{$停时$\tau: \tau < \infty$, a.s.$\}$，则存在$(\mathscr{F}_t)$可积的自然增过程$\alpha_{R_+}$使

$$\pi_t = E(\alpha_\infty | \mathscr{F}_t) - \alpha_t, \quad t \geqslant 0, [P], \tag{29.1}$$

其中 $\alpha_\infty = \lim \alpha_t$. 上述分解在随机等价意义下唯一.

系 29.4.8 设 X_{R_+} 为 $(\mathscr{F}_t)$ 右连续上鞅，且 $\{X_\tau, \tau \in \mathscr{T}\}$ 一致可积，则存在 $(\mathscr{F}_t)$ 右连续一致可积鞅 M_{R_+} 和 $(\mathscr{F}_t)$ 可积的自然增过程 α_{R_+}，使 X_{R_+} 有随机等价意义下唯一的分解：

$$X_t = M_t - \alpha_t, \quad t \geqslant 0, [P]. \tag{29.2}$$

T 离散时，增过程 (α_n) 称为自然的，如 α_n 为 $\mathscr{F}_{n-1}$ 可测. 任何 $(\mathscr{F}_n)$ 上鞅，在随机等价意义下有唯一分解 $X_n = M_n - \alpha_n$，$n \geqslant 0$，这里 M_n 为 $(\mathscr{F}_n)$ 鞅，而 α_n 为自然增过程. 此时可按如下递推方法直接写出这个分解：

令 $M_0 = X_0$，$\alpha_0 = 0$，对 $n \geqslant 1$，

$$M_n = M_{n-1} + [X_n - E(X_n | \mathscr{F}_{n-1})],$$
$$\alpha_n = \alpha_{n-1} + [X_{n-1} - E(X_{n-1} | \mathscr{F}_{n-1})], \quad n = 1, 2, \cdots.$$

可知 α_n 为自然增过程. 实由 $M_n = \sum_{i=1}^{n} (X_i - E(X_i | \mathscr{F}_{i-1}))$ 及 $\alpha_n = \sum^{n-1} (X_i - E(X_i | \mathscr{F}_{i-1}))$，知 $\alpha_n = M_n - X_n$，且关于 $\mathscr{F}_{n-1}$ 可测，此即为离散情形的可料性.

29.5 平方可积鞅

迈耶(Meyer)发现，对平方可积鞅，可像对布朗运动定义伊藤积分那样，来定义随机积分，从而为建立现代鞅论与随机积分论打开了通道. 因此平方可积鞅是一类重要的鞅.

本节恒设 $(\Omega, \mathscr{F}, P)$ 完备，$(\mathscr{F}_t)$ 右连续，一切鞅轨道右连续且有左极限. 又，对随机等价的两过程将不作区别，文中常不再指明.

29.5.1 定义与收敛性

定义 29.5.1 零初值 $(\mathscr{F}_t)$ 鞅 X_{R_+} 称为 $(\mathscr{F}_t)$ **平方可积鞅**

(square integrable martingale)，如∀ $t\geqslant 0$ 有 $EX_t^2<\infty$；称为($\mathscr{F}_t$)**一致**(uniformly)**平方可积鞅**，如$\sup\limits_{t\geqslant 0}EX_t^2<\infty$.

平方可积鞅全体记为 $\mathscr{M}_2$，一致平方可积鞅全体记为 $\mathfrak{M}_2$. 当又要求过程连续时，加记为 $\mathscr{M}_2^c$，$\mathfrak{M}_2^c$. 公式 29.2.3.(2)现在可写为：设 $M_{R_+}\in\mathfrak{M}_2$，则∀ $\lambda>0$，有

$$P(\sup|M_t|\geqslant\lambda)\leqslant\frac{1}{\lambda^2}\sup EM_t^2<\infty. \qquad (29.3)$$

例 29.5.2 一致平方可积鞅

(1) 设$[0,T]$时段内的独立增量过程(X_t)为可积[一致平方可积]，且 $EX_t=c$[$EX_t^2=c$]，c 为常数. 则 $X_{[0,T]}$为鞅[一致平方可积鞅]. 此命题由例 29.1.2 及定义 29.5.1，令 $\xi_{t_i}=X_{t_i}-X_{t_{i-1}}$，$i=1,2,\cdots,n$，$0=t_0<t_1<\cdots<t_n=t$ 可立即得到. 特别，零初值布朗运动$(B_t)_{0\leqslant t\leqslant T}$是($\mathscr{N}_t$)平方可积鞅，限 $t\in[0,T]$，则尚为一致的，这里 $\mathscr{N}_t=\sigma\{B_s:0\leqslant s\leqslant t\}$.

(2) 设 $a_t\in\mathscr{N}_t$，$E\int_0^T a_t^2\mathrm{d}t<\infty$，∀ $t\in[0,T]$，则可分过程 $\xi_t=\int_0^t a_s\mathrm{d}B_s$，$0\leqslant t\leqslant T$，$\in\mathfrak{M}_2$(对($\mathscr{N}_t$)，同(1)).

定理 29.5.3 对 $X\in\mathscr{M}_2$，令

$$\|X\|_{\mathscr{M}_2}=\sum_1^\infty((EX_n^2)^{\frac{1}{2}}\wedge 1)/2^n, \qquad (29.4)$$

则 (1) $\mathscr{M}_2$ 在准范数 $\|\cdot\|_{\mathscr{M}_2}$下成为弗雷歇(Frechet)空间(即完备可分距离线性空间，且有不变距离：$\rho(X^{(1)}+X,X^{(2)}+X)=\rho(X^{(1)},X^{(2)})$，$X$、$X^{(i)}\in\mathscr{M}_2$.

(2) $\mathscr{M}_2^c$ 是 $\mathscr{M}_2$ 的闭子空间.

定理 29.5.4 设 $M=(M_t)_{t\geqslant 0}$是右连续一致可积鞅，令 $M_\infty=\lim M_t$ a.s. 则

(1) $M\in\mathfrak{M}_2\Leftrightarrow EM_\infty^2<\infty$；

(2) $EM_\infty^2 = \sup\limits_t EM_t^2$，如 $M \in \mathfrak{M}_2$.

定理 29.5.5 记 $M_\infty = \lim M_t$ a.s.，$N_\infty = \lim N_t$，$M, N \in \mathfrak{M}_2$，

(1) $\mathfrak{M}_2$ 在内积$(M,N) = EM_\infty N_\infty$下为希尔伯特(Hilbert)空间，且与 $L^2(\Omega, \mathscr{F}_{\infty-}, P)$同构，映射 $M_1 \to M_\infty$ 为其同构映象；

(2) $\mathfrak{M}_2^c$ 为 $\mathfrak{M}_2$ 的闭子空间；

(3) 在 $\mathfrak{M}_2$ 上，$\|\cdot\|_{\mathfrak{M}_2}$强于 $\|\cdot\|_{\mathscr{A}_2}$.

29.5.2 正交性

定义 29.5.6 设 $M, N \in \mathfrak{M}_2$，且对$\forall$ 停时 τ，$EM_\tau N_\tau = 0$，则称鞅 M, N 相互正交，记为 $M \perp\!\!\!\perp N$.

注意 $\mathfrak{M}_2$ 定义中有零初值的约定，否则正交定义中应增 $M_0 N_0 = 0$.

定理 29.5.7 设 $M, N \in \mathfrak{M}_2$ 则 $M \perp\!\!\!\perp N \Leftrightarrow MN$ 为鞅.

定义 29.5.8 如(1) 对$\forall$ 停时 τ 及 $M \in \mathscr{H} \Rightarrow M^\tau \in \mathscr{H}$.

(2) 对$\forall$ $A \in \mathscr{F}_0$，$M \in \mathscr{H} \Rightarrow I_A M \in \mathscr{H}$.

称子空间 $\mathscr{H} \subset \mathfrak{M}_2$ 为稳定的(stable).

定理 29.5.9 设 $\mathscr{H}$ 为 $\mathfrak{M}_2$ 的稳定子空间，则 $\mathscr{H}^\perp \triangleq \{M \in \mathfrak{M}_2 : \forall\ N \in \mathscr{H}, M \perp\!\!\!\perp N\}$及 $\mathscr{H}$ 生成的闭线性子空间 $\overline{\mathscr{L}(\overline{\mathscr{H}})}$都是稳定的，且 $\mathscr{H}^\perp \perp\!\!\!\perp \overline{\mathscr{L}(\overline{\mathscr{H}})}$，这里 $M \perp N$，指 M, N 作为希尔伯特空间 $\mathfrak{M}_2$ 中两个元，为正交，即内积$(M,N) = EM_\infty N_\infty = 0$.

由此可证 $\mathfrak{M}_2^c$ 与$(\mathfrak{M}_2^c)^\perp \triangleq \mathfrak{M}_2^d$ 都是稳定的，且 $\mathfrak{M}_2^c \perp\!\!\!\perp \mathfrak{M}_2^d$，从而可建立：

定理 29.5.10 $\forall$ $M \in \mathfrak{M}_2$，有如下唯一分解：

$$M = M^c + M^d, \tag{29.5}$$

其中 $M^c \in \mathfrak{M}_2^c$，$M^d \in \mathfrak{M}_2^d$.

分别称(29.5)中 M^c，M^d 为 M 的连续鞅(continuous

martingale)部分和纯断鞅(purely discontinuous martingale)部分.

29.5.3 与$\mathfrak{M}_2$联系的增过程

定理 29.5.11 设$M \in \mathfrak{M}_2$,则M^2可作如下唯一分解:

$$M_t^2 = \mathfrak{M}_t + \langle M \rangle_t, t \geq 0, \tag{29.6}$$

其中$\mathfrak{M}_t$为零初值一致可积$(\mathscr{F}_t)$鞅,$\alpha_t = \langle M \rangle_t, t \geq 0$,是对应于$M$的可料增过程.

定义 29.5.12 称式(29.6)中$\langle M \rangle$为一致平方可积鞅M的特征. 又如$N \in \mathfrak{M}_2$,则M与N的**相互特征**(crossing character)$\langle M, N \rangle$由下面关系式定义

$$\langle M, N \rangle_t = \frac{1}{2}[\langle M + N \rangle_t - \langle M \rangle_t - \langle N \rangle_t]. \tag{29.7}$$

例 29.5.13 例 29.5.2(1)中布朗运动(B_t)的鞅特征$\langle B \rangle_t = t$,而例 29.5.2(2) 中过程(ξ_t) 的特征$\langle \xi \rangle_t = \int_0^t a_s^2 \mathrm{d}s$.

定理29.5.14 设$M, N \in \mathfrak{M}_2$,则

(1) M, N为正交增量过程.

(2) $E[(M_t - M_s)^2 | \mathscr{F}_s] = E[\langle M \rangle_t - \langle M \rangle_s | \mathscr{F}_s]$, $0 \leqslant s \leqslant t$.

(3) $M \perp\!\!\!\perp N \Leftrightarrow \langle M + N \rangle_t \Leftrightarrow \langle M \rangle_t + \langle N \rangle_t \Leftrightarrow \langle M, N \rangle = 0$.

29.6 局部鞅、鞅刻画与鞅表现

设概率空间$(\Omega, \mathscr{F}, P)$完备,$(\mathscr{F}_t)$完备且右连续.

29.6.1 局部鞅定义与性质

定义29.6.1 两增过程(见定义29.4.4)之差,称为有限变差过程(process of finite variation). 设A为增过程,且$A_\infty = \lim A_t$

为可积 r.v.，则称 A 为**可积增过程**. 可积的有限变差过程为**可积变差过程**. 易"过程"为"鞅"，则有**变差鞅**(martingale of finite variation)及**可积变差鞅**(martingale of integrable variation)等概念.

定义29.6.2 右连续适应过程 M 如关于$(\mathscr{F}_t)$存在停时列 $\tau_n \uparrow \infty$ a.s.，使每个 $M^{\tau_n}-M_0 \triangleq \{M_{\tau_n \wedge t}-M_0, t \geqslant 0\}$ 为$(\mathscr{F}_t)$一致可积鞅(可积变差鞅)，则称为**局部鞅**(local martingale)[**局部可积变差鞅**(local martingale of integrable variation)].

性质29.6.3 局部鞅性质

(1) 设 M 为局部鞅，τ 为停时，则 M^τ 为局部鞅.

(2) 设 M 为非负局部上鞅(仿(29.6.2)定义)，M_0可积，则 M 是上鞅.

(3) M 为局部鞅 $\Leftrightarrow$ $\exists$ 停时 $\tau_n \uparrow \infty$ a.s.，使 $M^{\tau_n} I_{(\tau_n > 0)}$ 为一致可积鞅.

(4) 设 τ_n 为停时，M^{τ_n}为局部鞅，$n=1,2,\cdots$且 $\sup \tau_n = \infty$ a.s.，则 M 为局部鞅.

(5) M 为局部鞅，τ 为停时，实值 $\xi \in \mathscr{F}_\tau$，则 $\xi(M-M^\tau)$为局部鞅.

性质29.6.4 局部鞅的分解 局部鞅概念最早引入于下一分解.

(1) 伊藤-渡边(Ito-Watanabe)分解 设 X 为非负右连续上鞅，则 X 有唯一分解 $X=M-A$，其中 M 为 L_1中有界局部鞅，A 为$(\mathscr{F}_t)$可积可料(29.6.5)增过程.

(2) 局部鞅基本定理 设 M 为局部鞅，则对$\forall\ \epsilon>0$，可作如下分解：$M=M_0+U+V$，其中 U 为零初值有界局部鞅，$|\Delta U|<\epsilon$，V 为局部可积变差鞅.

(3) 局部鞅分解定理 局部鞅 M 有唯一分解 $M=M^c+M^{du}+M^{di}$，其中 M^c 为连续局部鞅，M^{du}为只有可及跳的纯断局部鞅(定

义29.6.7)，M^{di}为只有不可及跳(定义29.6.6)的纯断局部鞅.

定义29.6.5 当由关于 $R_1\times\Omega$ 上全体左连续($\mathscr{F}_t$)适应过程所产生的 σ 代数 $\mathscr{P}$ 为可测的过程时，称它为**可料过程**(predictable process).

定义29.6.6 如$(\alpha<t)\in\mathscr{P},\forall\ t<\infty$. 称 $R_+\times\Omega$ 上非负函数 α 为可料时(predutable time). 当 τ 为停时，$\exists$ 可料时列 α_n 使 $P(\exists\ n,\alpha_n=\tau<\infty)=P(\tau<\infty)$，则称 τ 为 **a.s. 可及时**(accessible time). 如对一切可料时 α，$P(\tau=\alpha<\infty)=0$，则称 τ 为**不可及时**. 如$(A_t)_{t\geqslant0}$的一切跳跃时为 a.s. 可及时[不可及时]，则称(A_t)只有**可及跳过程**(accessible jump process)[**不可及跳过程**].

定义29.6.7 如 $M_0=0$，且$\exists$ 零初值局部纯断平方可积鞅 U 及零初值有限变差局部鞅 V，使 $M=U+V$. 称局部鞅 M 为纯断局部鞅(purely discontinue local martingale).

29.6.2 布朗运动的鞅刻画与鞅表现

定理29.6.8 列维鞅刻画 设 B 为零初值连续适应过程，则 B 为($\mathscr{F}_t$)BM 中取 $d=1\Leftrightarrow B$ 为局部鞅，且其特征$\langle B\rangle_t=t$.

列维鞅刻画的一个重要应用是

定理29.6.9 设 M 为零初值($\mathscr{F}_t$)连续局部鞅，$\lim\limits_{t\to\infty}\langle M\rangle_t=\infty$，a.s. 令 $\tau_t=\inf(s:\langle M\rangle_s>t)$，则$(M_{\tau_t})$为$(\mathscr{F}_{\tau_t})$BM.

定理29.6.10 BM 的鞅表现 ($\mathscr{F}_t$)BM $B\triangleq(B_t)_{t\geqslant0}$. 关于($\mathscr{Y}_t$)有可料表示性，即对任何($\mathscr{Y}_t$)局部鞅 N，$N_0=0$，可表为一可料过程 H 对($\mathscr{Y}_t$)BM 的随机积分，这里 $\mathscr{Y}_t=\sigma\{\mathscr{N}_t^0,\mathscr{N}\}$，$\mathscr{N}_t^0=\sigma\{B_s,s\leqslant t\}$，$\mathscr{N}=\{A:\in\mathscr{F},\mathrm{P}(A)=0\}$.

当 B 为($\mathscr{F}_t$)BM 时，B 必为($\mathscr{Y}_t$)BM，且对每 $t\in R_+$，$\mathscr{Y}_+=\mathscr{Y}_t=\mathscr{Y}_-$. 又，$Y_t\triangleq B_t^2-\sigma^2t$，$Z_t\triangleq\exp\{\theta B_t-\frac{1}{2}\theta^2\sigma^2t\}$，$\theta\in R_1$，都是鞅. 反之，如 X_t、$X_t^2-\sigma^2t$ 都是鞅的($\mathscr{F}_t$)过程 X_T，必为 BM.

29.6.3 泊松过程的刻画与表现

定义29.6.11 设(P_t)为纯断适应增过程,$P_0=0$,其跳为$+1$.又如对一切$0\leqslant s<t<\infty$,P_t-P_s与$\mathscr{F}_s$独立,且$P_t-P_s\sim P(t-s)$,即$P(P_t-P_s=k)=(t-s)^k e^{-(t-s)}/k!$,$k=0,1,\cdots$则称$(P_t)$为泊松过程.

定理29.6.12 泊松过程鞅刻画 设(P_t)为纯断适应增过程,$P_0=0$,跳为1,则(P_t)为泊松过程$\Leftrightarrow(P_t-t)$为鞅$\Leftrightarrow(P_t-t)$为局部鞅.

定理29.6.13 泊松过程鞅表示 设(P_t)为泊松过程,$Q_t=P_t-t$, $t\geqslant 0$.则一切零初值$(\mathscr{Y}_t)$局部鞅,可表为一可料过程对(Q_t)按轨道的斯蒂尔切斯积分.

定理29.6.14 设$P_t\sim P(\lambda t)$,$\lambda>0$,则$Y_t\triangleq P_t-\lambda t$及$V_t\triangleq\exp(-\theta P_t+\lambda t(1-e^{-\theta}))$,$\theta\in R_1$,都是鞅.

29.7 应用

本节除继续介绍鞅的应用背景之外,还介绍如何构造鞅及应用鞅的几个结论.

例29.7.1 随机游动问题 考虑例29.1.7中一维游动,$X_n=(q/p)^{S_n}$为鞅.令$h_x=\inf\{n:S_n=x\}$,x为整数,$\nu_a=P(h_{-a}<h_b)$,即(S_n)在到达b之前先到$-a$的概率.对鞅X_n用(29.2.4(1))并取$\tau=h_{-R}\wedge h_b\triangleq\min\{h_{-a},h_b\}$,它是停时,则$1=EX_0=EX_\tau=r^{-a}\nu_a+r^b(1-\nu_a)$,这里$r=q/p$.从而

$$\nu_a=(1-r^b)/(r^{-a}-r^b).$$

利用(29.2.4(2)),可建立在概率统计中广为应用的沃尔德(Wald)公式和沃尔德基本公式.

公式29.7.2 沃尔德公式 设(ξ_n)为i.i.d列,$E|\xi_n|<\infty$,τ

为停时，且 $E\tau<\infty$，则

$$E\left(\sum^{\tau}\xi_i\right)=E\xi_1 E\tau. \tag{29.8}$$

公式 29.7.3 沃尔德基本公式（Wald's fundamental identity） 设(ξ_n)为 i.i.d 列，且非退化，$\xi_0=0$. $\tau=h_{(-\infty,-a)}\wedge h_{[b,\infty)}$，即首次跑出$[-a,b)$的时间，它是停时. 又设 $\Phi(\lambda)$为 ξ_1 的母函数，即 $\Phi(\lambda)=Ee^{\lambda\xi_1}$，$\lambda>0$. 则

$$E(\Phi(\lambda)^{-\tau}e^{\lambda S_\tau})=1, \tag{29.9}$$

其中 λ 满足 $\Phi(\lambda)\geqslant 1$，而 $S_n=\sum_1^n\xi_i$.

证 令 $X_0=1$，$X_n=\Phi(\lambda)^{-n}e^{\lambda S_n}$，$n\geqslant 1$. 由(29.1.7)易知 S_n 对(ξ_n)为鞅. 这一结论也可直接证明如下：

$$\begin{aligned}E(X_{n+1}|\xi_0,\cdots,\xi_n)&=E[\Phi(\lambda)^{-n-1}e^{\lambda\sum_1^n\xi_i}e^{\lambda\xi_{n+1}}|\xi_0,\cdots,\xi_n)\\&=\Phi(\lambda)^{-n-1}e^{\lambda\sum_1^n\xi_i}E(e^{\lambda\xi_{n+1}})\end{aligned}$$

（这里用到$\exp(\lambda\sum_1^n\xi_i)\in\sigma\{\xi_0,\cdots,\xi_n\}$、条件期望性质及 ξ_{n+1} 与 $\xi_0,\cdots,\xi_n$ 独立）

$$=\Phi(\lambda)^{-n}\exp\left(\lambda\sum_1^n\xi_i\right)=X_n.$$

现在验证对 X_n，定理29.2.4(2)条件满足

$$\begin{aligned}E(|X_{n+1}-X_n||\xi_0,\cdots,\xi_n|)&=X_nE(|\Phi(\lambda)^{-1}e^{\lambda\xi_{n+1}}-1|)\\&\leqslant 2X_n\leqslant 2e^{\lambda S_n}.\end{aligned}$$

故 $n<\tau$ 时，$S_n<b$. 从而前式右方$\leqslant 2e^{\lambda b}$.

又由 ξ_1 非退化，对 $c=a+b$，$\exists$ N 及 $\delta>0$，使 $P(|\xi_N|>c)>\delta$. 故

$$\begin{aligned}P(\tau\geqslant kN)&\leqslant P(|S_N|\leqslant c)\cdot P(|S_{2N}-S_N|\leqslant c)\\&\quad\cdots P(|S_{kN}-S_{kN-N}|\leqslant c)\leqslant(1-\delta)^k,\end{aligned}$$

且 $P(\tau\geqslant n)\downarrow$，从而 $E\tau<\infty$.

对 X_n 用定理29.2.4(2)，$1=EX_0=EX_\tau=E(\Phi(\lambda)^{-\tau}e^{\lambda S_\tau})$. 式

(29.9)得证.

将式(29.9)对 λ 形式求导,可推得式(29.8).

例29.7.4 似然比序列(the sequence of likelihood ratios)

设$\xi_n, n\geqslant 0$,为i.i.d,而f_1、f_0为两密度函数,且 $f_0>0$. 令$X_n=\prod_0^n (f_1(\xi_i)/f_0(\xi_i)), n\geqslant 0$,则当 ξ_i 的密度就是 f_0 时,X_n 是鞅. 事实上由 $Ef_1(\xi_{n+1})/f_0(\xi_{n+1}) = \int f_1(y)\Big/ f_0(y)f_0(y)\mathrm{d}y = \int f_1(y)\mathrm{d}y = 1$,而 $E(X_{n+1} \mid \xi_0,\cdots,\xi_n) = E(X_n f_1(\xi_{n+1})/f_0(\xi_{n+1}) \mid \xi_0,\cdots,\xi_n) = X_n E(f_1(\xi_{n+1})/f_0(\xi_{n+1})) = X_n$, 得证.

似然比序列构造的鞅,在序贯检验各步骤性能的评价中有许多用处.

取 f_0 为 $N(0,\sigma^2)$的密度,f_1为 $N(\mu,\sigma^2)$密度时,得鞅 $X_n = \exp\left(\frac{\mu}{\sigma^2}\sum_1^n \xi_i - \frac{n\mu^2}{2\sigma^2}\right), n\geqslant 1$.

例29.7.5 水库模型中的应用 令 Z_t 为容量b的水库在t时的水位,I_t,O_t 分别是$(t,t+1]$时段内随机流入和流出的水量. 这时有如下平衡方程

$$Z_{t+1} = \min((Z_t + I_t - O_t)^+, b).$$

设水库水位应维持在水平 a 之上,以应付航行、发电等需要. 令 $\tau=h_{(-\infty,a]}$,则希望 $E\tau$ 尽量大. 我们来给出 $E\tau$ 的一个下界,对设计与规划均有一定意义.

令净流入 $\xi_{t+1}=I_{t+1}-O_{t+1}$,并设

$$E(\xi_{t+1} \mid \xi_1,\cdots,\xi_t) \leqslant m, E(\mathrm{e}^{-\lambda\xi_{t+1}} \mid \xi_1,\cdots,\xi_t) \leqslant 1, \tag{29.10}$$

其中 m,λ 为已知正常数. 又令

$$f(z) = \begin{cases} \frac{1}{m}(\mathrm{e}^{\lambda(b-a)}(1-\mathrm{e}^{-\lambda(z-a)})/\lambda - (z-a)), & a\leqslant z\leqslant b, \\ 0, & z<a, \\ f(b), & z>b. \end{cases}$$

这里 $z=Z_0$.

先证 $X_n \triangleq f(Z_n)+n$ 对 $\{\xi_n\}$ 为下鞅. 令

$$g(z)=\frac{1}{m}(\lambda^{-1}e^{\lambda(b-a)}(1-e^{-\lambda(z-a)})-(z-a)), \text{一切 } z\in R_1,$$

它在 $z=b$ 之前上升而其后下降, 且 $f\geqslant g$.

$$\begin{aligned}E(X_{n+1}|\xi_0,\cdots,\xi_n) &= E(f(Z_n+\xi_{n+1})|\xi_0,\cdots,\xi_n)+(n+1)\\ &\geqslant E(g(Z_n+\xi_{n+1})|\xi_0,\cdots,\xi_n)+(n+1).\end{aligned}$$

令 U 为 r. v., 且 $EU\leqslant m, Ee^{-\lambda U}\leqslant 1$, 则

$$\begin{aligned}Eg(z+U) &= m^{-1}(\lambda^{-1}e^{\lambda(b-a)}(1-Ee^{-\lambda(z+U-a)})-(z+EU-a))\\ &\geqslant f(z)-1, \quad a\leqslant z\leqslant b.\end{aligned}$$

注意式(29.10), 可得 $E(X_{n+1}|\xi_0,\cdots,\xi_n)\geqslant f(Z_n)+n=X_n$.

下证 $E\tau\geqslant f(z)$.

由定理29.2.4(4)知

$$E(f(Z_{\tau\wedge n})+(\tau\wedge n))\geqslant E(f(Z_0))=f(z).$$

注意 f 有界及 $Z_\tau\leqslant a$, 令 $n\to\infty$,

$$f(z)\leqslant \lim_{n\to\infty}E(f(Z_{\tau\wedge n})+(\tau\wedge n))=E(f(Z_\tau))+E\tau=E\tau.$$

当 $\xi_i\sim N(\mu_i,\sigma_i^2)$, 且彼此独立, $\mu_i\leqslant m$, $\sigma_i^2\leqslant 2\mu_i/\lambda$ 时, 条件(29.10)一定满足.

30 随机微分方程

30.1 引言

近年来，随机微分方程在系统科学、工程控制、经济管理与金融及生态学等诸多方面有广泛应用. 随机微分方程是研究一般马尔可夫过程的一种有效工具：借助于简单的随机过程造出一般过程的轨道，它与扩散过程有密切联系.

例 30.1.1 考虑一悬浮在液体中的粒子运动，它主要可分为两部分：非随机(即确定性)运动(由液体运动和附加外力引起)和随机运动(由粒子间碰撞与相互作用引起). 碰撞产生的独立位移，依中心极限定理知可近似视为正态的量，因此后一随机运动常可用 BM(布朗运动)描述. 这样在 t 到 $t+\Delta t$ 时段沿 x 轴的位移可近似为

$$X_{t+\Delta t} - X_t \approx \mu(t,x)\Delta t + \sigma(t,x)\Delta B_t, \qquad (30.1)$$

这里 X_t 是在时刻 t 该粒子的位置. μ 是粒子在时刻 t 和位置 x 处的瞬时速度，$\Delta B_t = B_{t-\Delta t} - B_t$ 是 BM 的增量，而 $\sigma^2(t,x)>0$ 是与过程 $X \triangleq (X_t)$ 的碰撞有关的瞬时方差. 又设 $X_0=0, X_t=x$. (30.1)右方第一项反映决定力引起的运动，而第二项表达运动的随机部分.

由于 $B \triangleq (B_t)$ 有连续轨道，对 μ, σ^2 加上一定条件可希望 X 为连续的，从而是一扩散过程，其无穷小漂移系数与扩散系数分别为(注意定义 26.1.2 及 $E(\Delta X_t)^2 = O(\Delta t)^2$)

$$\lim_{\Delta t \downarrow 0} E\Delta X_t/\Delta t = \mu(t,x),$$

$$\lim_{\Delta t \downarrow 0} E(\Delta X_t)^2/\Delta t = \lim[V(\sigma(t,x)\Delta B_t)/\Delta t$$

$$+ (E\Delta X_t)^2]/\Delta t = \sigma^2(t,x).$$

BM 轨道以概率 1 处处连续，却处处不可微，故必须赋予"dB_t"一种新的意义，从而由(30.1)经极限形式得到

$$dX_t = \mu(t,x)dt + \sigma(t,x)dB_t, \tag{30.2}$$

或

$$X_t = X_0 + \int_0^t \mu(t,X_s)ds + \int_0^t \sigma(s,X_s)dB_s. \tag{30.3}$$

例 30.1.2 物理学中著名的郎之万(Langevin)方程是又一例子：设 α,β 为常数，有

$$dX_t = -\alpha X_t dt + \beta dB_t.$$

其物理意义及方程的解见节 30.5.2.

伊藤首先对 dB_t 给出了定义，以(30.3)表达扩散过程的解. 这样，用简单的 BM 直接给出一个扩散过程轨道 X_t，而不是只造出 X 的某个特征如转移函数(科尔莫戈罗夫构造法)或无穷小算子(费勒半群方法). 更进一步，对其他简单过程或随机测度定义微分与积分，则在合适的条件下，仿上可直接得到更一般过程的轨道. 这种构造过程的办法，由一维推广到多维，也不会产生本质的困难.

注 30.1.3 记号 $SD[E]$：伊藤随机微分[方程]，$SI[E]$：伊藤随机积分[方程].

30.2 对布朗运动的随机积分

30.2.1 对布朗运动的伊藤积分

设 $B=\{B_t,t\in[a,b]\}$ 是可分$(\mathscr{F}_t)$适应的布朗运动，且对每一 $t\in[a,b]\triangleq\Delta$，$\mathscr{F}_t$ 与 $\mathscr{N}_b^t(B)\triangleq\sigma\{B_s-B_t,s\in[t,b]\}$独立，即如 $A_i\in\mathscr{F}_t$，$B_j\in\mathscr{N}_b^t(B)$，$i=\overline{1,n}$，$j=\overline{1,m}$. 则

$$P(A_iB_j,i=\overline{1,n},j=\overline{1,m}) = \prod_{i=1}^n P(A_i)\prod_{j=1}^m P(B_j).$$

本节目的是满足一定条件的二元函数 $f(t,\omega)$ 定义伊藤积分

$$I(f,\omega)=\int_a^b f(s,\omega)\mathrm{d}B_s(\omega)\triangleq\int_a^b f_s\mathrm{d}B_s. \qquad (30.4)$$

定义 30.2.1 简单函数(simple function)满足下列条件的实值函数 $f(t,\omega)$，称为简单函数.

(1) f 为二元 $\mathscr{B}_{[a,b]}X\mathscr{F}_t$ 可测;

(2) 固定 $t\in\Delta=[a,b]$，f 为 $\mathscr{F}_t$ 可测;

(3) $\exists$ 分割：$a=a_0<a_1<\cdots<a_n=b$，a_i 不依赖于 ω，使

$$f(t,\omega)=f(a_{i-1},\omega),a_{i-1}\leqslant t<a_i,\quad i=\overline{1,n}$$

且 $\qquad f(a_i,\cdot)\in L^2[\Omega]\triangleq\{\varphi\cdot\int_\Omega|\varphi(\omega)|^2P(\mathrm{d}\omega)<\infty\}.$

注 30.2.2 记号

S：简单函数全体.

M_0： 满足定义 30.2.1(1)及 30.2.1(2)的函数全体.

M： $\int_a^b|f(t,\omega)|^2\mathrm{d}t<\infty$，a.s. 的 M_0 中的函数全体.

$\overline{S}$： $\int_a^b E|f(t,\omega)|^2\mathrm{d}t<\infty$ 的 M_0 中的函数全体.

定义 30.2.3 对 BM 的伊藤积分(Ito integral for a Brownian motion) 设 $f\in M$，如下定义(30.4)中之 $I(f,\omega)$：

(1) 如 $f\in S$，令

$$I(f,\omega)=\sum_1^n f(a_{i-1},\omega)[B_{a_i}(\omega)-B_{a_{i-1}}(\omega)].$$

(2) 如 $f\in\overline{S}$，$\{f_n\}\subset S$，且 $f_n\xrightarrow{L_2}f$，令

$$I(f,\omega)=\text{l.i.m}\quad I(f_n,\omega),\text{a.s. P}.$$

(3) 如 $f\in M$，令

$$f_N(t,\omega)=f(t,\omega)I_{[-N,N]}\left(\int_a^t|f(s,\omega)|^2\mathrm{d}s\right),$$

$$I(f,\omega)\overset{P}{=}\lim_{N\to\infty}I(f_N,\omega).$$

可证 S 在 $\overline{S}$ 中稠(依 $L^2(\Delta\times\Omega)$ 中范数 $\|f\|^2=\int_a^b E|f(t,\omega)|^2\mathrm{d}t$),且定义 30.2.3(2) 中 $I(f,\omega)$ 惟一. 对 $\forall\ \{f_n\}\subset\overline{S}$, $\int_a^b|f-f_n|^2\mathrm{d}t\xrightarrow{P}0$, 有 $I(f_n,\omega)\xrightarrow{P}I(f,\omega)$.

30.2.2 伊藤积分的性质

设 $f,g\in M$, 令 $\int_A f_t\mathrm{d}B_t=I(fI_A,\omega)$, $A\in\mathscr{B}_\Delta$, 这里 I_A 为 A 之示性函数.

性质 30.2.4 基本性质

(1) $I(f,\omega)=\int_a^c f_t\mathrm{d}B_t+\int_c^b f_t\mathrm{d}B_t,\quad a\leqslant c\leqslant b.$

(2) 线性 $I(\alpha f+\beta g,\omega)=\alpha I(f,\omega)+\beta I(g,\omega)$.

(3) $I(I_{[t_1,t_2]},\omega)=B_{t_2}-B_{t_1}$, $t_1,t_2(>t_1)\in\Delta$.

(4) 如 $f,g\in\overline{S}$, 则 $EI(f,\omega)=0$. $E(I(f,\omega)I(g,\omega))=\int_a^b Efg\mathrm{d}t$,

特别地,
$$E(I(f,\omega))^2=\int_a^b E(f_t)^2\mathrm{d}t.$$

(5) $P(|I(f,\omega)|>0)\leqslant P(\sup\limits_{t\in\Delta}|f_t|>0)$,

$$P(\sup_{t\in\Delta}|I(fI_{[a,t]},\omega)|>0)\leqslant P(\sup_{t\in\Delta}|f_t|>0).$$

定理 30.2.5 矩估计(estimations of moments) 设 $f\in M$, 且对某 $p>0$, $E\left(\int_\Delta|f_t|^2\mathrm{d}t\right)^{p/2}<\infty$, 则存在常数 α_p 和 β_p, 使

$$E\left|\int_\Delta f_t\mathrm{d}B_t\right|^p\geqslant\alpha_pE\left(\int_\Delta|f_t|^2\mathrm{d}t\right)^{p/2},\quad 如\ p>1;$$

$$\leqslant\beta_pE\left(\int_\Delta|f_t|^2\mathrm{d}t\right)^{p/2},\quad 如\ p>0.$$

$p=2$ 时,两式均取等号,且 $\alpha_2=\beta_2=1$.

定理 30.2.6 设 $f\in M$, $\Delta=[0,b]$, $\xi_t\triangleq\int_0^t f_s\mathrm{d}B_s$, 且取过程

(ξ_t)为可分的，则(ξ_t)的样本函数以概率 1 连续，且对任意正数 c 和 N，有

$$P\left(\sup_{t\in\Delta}\left|\int_0^t f_s \mathrm{d}B_s\right| > c\right) \leqslant P\left(\int_0^b f_s^2 \mathrm{d}s > N\right) + N/c^2. \quad (30.5)$$

定理 30.2.7 设 $f\in\overline{S}$，$\Delta=[0,b]$，则过程 $\zeta_t \triangleq \int_0^t f_s \mathrm{d}B_s, t\in\Delta$ 是连续的一致平方可积鞅，其特征为 $\langle\zeta\rangle_t = \int_0^t f_s^2 \mathrm{d}s$. 这时

$$P(\sup_{t\in\Delta}|\zeta_t| > c) \leqslant \frac{1}{c^2}E\int_0^b f_s^2 \mathrm{d}s,$$

$$E(\sup_{t\in\Delta}|\zeta_t|^2) \leqslant 4E\int_0^b f_s^2 \mathrm{d}s. \quad (30.6)$$

30.2.3 对布朗运动的 S 积分

许多生物、物理模型作扩散近似时，其极限过程归结为对布朗运动的斯特拉托诺维奇(Stratonovich)积分(简称 S 积分或 C 积分)意义下的 SIE，解得的扩散过程，比伊藤积分意义下的更自然：对输入过程(X_t)的样本轨道的直接积分，其复合过程的 S 随机微分，有通常的全微分公式(30.13). 只是这种 S 积分存在的条件较 I 积分更为苛刻，它也不能像后者生成鞅.

定义 30.2.8 对 BM 的 S 积分(S integral for a Brownian motion) 设 $f(t,\omega)$为$[a,b]\times\Omega$上二元函数，如对$[a,b]$的任一分割$\{a_i\}$：$a=a_0<a_1<\cdots<a_n=b$，只要 $n\to\infty$、$\Delta_n\triangleq\max_{1\leqslant i\leqslant n}|a_i-a_{i-1}|\to 0$ 时，下述均方极限

$$\underset{n\to\infty,\Delta_n\to 0}{\mathrm{l.i.m}}\sum_{i=1}^{n}\frac{1}{2}(f(a_{i-1},\omega)+f(a_i,\omega))[B_{a_i}(\omega)-B_{a_{i-1}}(\omega)] \quad (30.7)$$

存在，且有同一极限(对固定 $\omega\in\Omega$)，则称 f 对 BM(B_t)的 S 积分存在，其 S 积分值定义为上述极限值，并记为

$$S(f,\omega)=(S)\int_a^b f(t,\omega)\mathrm{d}B_t(\omega)\triangleq(S)\int_a^b f_t\mathrm{d}B_t. \quad (30.8)$$

特别，当 $f\in S$ 且 $\{a_i\}$ 为定义 30.2.1 中分割，则

$$S(f,\omega)=\sum_{i=1}^n\frac{1}{2}(f(a_{i-1},\omega)+f(a_i,\omega))[B_{a_i}(\omega)-B_{a_{i-1}}(\omega)]. \quad (30.9)$$

性质 30.2.9 伊藤积分与 S 积分的几个差别

（1）依定义可算得 BM 对 BM 的随机积分

$$\begin{cases}\int_a^b B_s\mathrm{d}B_s=\frac{1}{2}(B_b^2-B_a^2)-\frac{1}{2}(b-a),\\(S)\int_a^b B_s\mathrm{d}B_s=\frac{1}{2}(B_b^2-B_a^2).\end{cases} \quad (30.10)$$

可见对伊藤积分，一般的分部积分公式不成立. 伊藤积分 $\int_a^b B_s\mathrm{d}B_s$ 也可由公式 30.3.3 取 $\varphi(t,x)\equiv x^2$ 得到.

（2）SDE 解间的关系　仿定义 30.3.1 可定义 S 意义下的随机微分. 设 X_t 是 S 意义下 SDE,

$$\mathrm{d}X_t=a(t,X_t)\mathrm{d}t+b(t,X_t)\mathrm{d}B_t \quad (30.11)$$

的解，则它也是伊藤意义下的 SDE.

$$\mathrm{d}X_t=\left[a(t,X_t)+\frac{1}{2}b_x(t,X_t)b(t,X_t)\right]\mathrm{d}t+b(t,X_t)\mathrm{d}B_t \quad (30.12)$$

的解. 这里 $b_x(t,X_t)=\left[\frac{\partial}{\partial x}b(t,x)\right]_{x=X_t}$. 当 $b(t,x)$ 与 x 无关时，两种随机微分意义下的 SDE 的解一致.

（3）复合函数过程有通常的全微分公式（在 S 微分意义下），即若 X_t 是(30.11)的解（S 微分意义），则 $Y_t=f(t,X_t)$ 是 S 微分意义下,

$$\mathrm{d}Y_t=f_t(t,X_t)\mathrm{d}t+f_x(t,X_t)\mathrm{d}X_t \quad (30.13)$$

的解，只要 f 满足适当条件使上述解存在. (30.12)有通常的全微

分形式. 而在伊藤随机微分意义下的(30.10)一般无此性质.

(4) 伊藤意义下的不定随机积分 $\int_0^t f_s \mathrm{d}B_s$ 常确定一类重要的鞅. 如 $f_t \in \mathscr{N}_t \triangleq \sigma\{B_s; 0 \leqslant s \leqslant t\}, E\int_0^T f_t^2 \mathrm{d}t < \infty$, 则可分过程 $M_t = \int_0^t f_s \mathrm{d}B_s$, 在 $0 \leqslant t \leqslant T$, 为一致平方可积鞅. 而(S)$\int_0^t f_s \mathrm{d}B_s$ 一般却不是鞅.

(5) 在实际问题求解中的差异, 见节 30.5 及[70].

30.3 伊藤微分方程与积分方程

30.3.1 随机微分的定义与性质

设 B 是 $(\mathscr{F}_t)$BM, X、μ 及 $\sigma \in M_0$, 且$\int_a^b |\mu_t| \mathrm{d}t < \infty, \int_a^b |\sigma_t|^2 \mathrm{d}t < \infty$.

定义 30.3.1 随机微分(stochastic differential), 如果对 $a \leqslant s \leqslant t \leqslant b$, 有

$$X_t - X_s = \int_s^t \mu_\tau \mathrm{d}\tau + \int_s^t \sigma_\tau \mathrm{d}B_\tau, [P],$$

且 X_t 为 t 的连续函数$[P]$, 称随机过程 X 有随机微分(SD)$\mathrm{d}X_t$, 且 $\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t$.

性质 30.3.2 基本性质

(1) 随机微分运算是线性的.

(2) 如 X_i 有随机微分 $\mathrm{d}X_i(t) = \mu_i(t)\mathrm{d}t + \sigma_i(t)\mathrm{d}B(t), i = 1, 2$. 则积 $X_1(t)X_2(t)$ 有随机微分, 且

$$\mathrm{d}X_1(t)X_2(t) = X_1(t)\mathrm{d}X_2(t) + X_2(t)\mathrm{d}X_1(t) + \sigma_1(t)\sigma_2(t)\mathrm{d}t.$$

(3) 如 X 有 SD(定义 30.3.1), $\varphi(t, x)$ 为 $\Delta \times R^1$ 上连续函数, 且 ϕ_t, ϕ_x 及 ϕ''_{xx} 连续, 则复合过程 $\varphi(t, X_t)$ 有 SD, 且有

公式 30.3.3 伊藤变换公式(Ito transformation formulae)

$$d\varphi(t,X_t) = [\varphi'_t(t,X_t) + \varphi'_x(t,X_t)\mu_t + \frac{1}{2}\varphi''_{xx}(t,X_t)\sigma_t^2]dt$$
$$+ \varphi'_x(t,X_t)\sigma_t dB_t.$$

更为一般的伊藤变换公式是:设

$$dX_i(t) = \mu_i(t)dt + \sigma_i(t)dB(t), i = \overline{1,n}, f(t,x_1,\cdots,x_n)$$

连续且有连续偏导 φ'_t、φ'_{x_i}及 $\varphi''_{x_ix_j}, i,j=\overline{1,n}$,则 $\varphi(t,X_1,X_2,\cdots,X_n)$有 SD,且有

$$\begin{aligned} d\varphi(t,X_1,\cdots,X_n) = \Big[&\frac{\partial\varphi}{\partial t}(t,X_1(t),\cdots,X_n(t)) \\ &+ \sum_{i=1}^{n}\varphi'_{x_i}(t,X_1(t),\cdots,X_n(t))\mu_i(t) \\ &+ \frac{1}{2}\sum_{i,j=1}^{n}\varphi''_{x_ix_j}(t,X_1(t),\cdots,X_n(t))\sigma_i(t)\sigma_j(t)\Big]dt \\ &+ \sum_{i=1}^{n}\varphi'_{x_i}(t,X_1(t),\cdots,X_n(t))\sigma_i(t)dB(t). \end{aligned}$$

下一定理在一定条件下保证积分号下取极限,而定理 30.3.6 将 SDE 化为 SIE.

定理 30.3.4 设 $f_n, f\in M$,且 $\int_a^b |f_n(t) - f(t)|^2 dt \xrightarrow{P} 0$,则下列可分过程间有关系

$$\sup_{a\leqslant t\leqslant b}\left|\int_a^t f_n(s)dB_s - \int_a^t f(s)dB_s\right| \xrightarrow{P} 0.$$

即对 $t\in\Delta$,一致地有 $\int_a^t f_n dB_s \xrightarrow{P} \int_a^t f dB_s$.

定理 30.3.5 如 $dX_t = \mu_t dt + \sigma_t dB_t$,则

$$(X_t - X_s)^2 = \int_s^t [2(X_\tau - X_s)\mu_\tau + \sigma_\tau^2]d\tau + \int_s^t 2[X_\tau - X_s]\sigma_\tau dB_\tau.$$

定理 30.3.6 矩的恒等式(identity of moments) 设 $f\in M$,$\Delta=[0,b]$,且对某 n,$E\left(\int_\Delta f_t^2 dt\right)^{n/2} < \infty$,则对 $\forall\ \alpha,\beta\in R^1$,过程 $G_n\left(\alpha + \int_0^t f_s^2 ds, \beta + \int_0^t f_s dB_s\right)$为$(\mathscr{F}_t)$ 鞅,特别

$$EG_n\left(\alpha+\int_\Delta f_s^2\mathrm{d}s,\beta+\int_\Delta f_s\mathrm{d}B_s\right)=G_n(\alpha,\beta).\qquad(30.14)$$

这里 $G_n(t,x)=t^{n/2}He_n(t^{-1/2}x)$，$t\geqslant0$，$x\in R^m$，$n\geqslant0$，而 $He_n(z)$ 为埃尔米特(Hermite)多项式：

$$He_n(z)=(-1)^n\exp(z^2/2)\frac{\mathrm{d}^n}{\mathrm{d}z^n}\exp(-z^2/2).$$

前 5 个 $G_n(t,x)$：

$$G_0(t,x)=1,\quad G_1(t,x)=x,\quad G_2(t,x)=x^2-t,$$

$$G_3(t,x)=x^3-3tx,\quad G_4(t,x)=x^4-6x^2t+3t^2.$$

例 30.3.7 伊藤公式与矩恒等式的应用

(1) 试求 $\mathrm{d}Y_t$，其中 $Y_t=\mathrm{e}^{X_t}$，而

$$\mathrm{d}X_t=a_t\mathrm{d}t+b_t\mathrm{d}B_t.\qquad(30.15)$$

取 $\varphi(t,x)=\mathrm{e}^x$，则由伊藤公式 30.3.3

$$\begin{aligned}\mathrm{d}Y_t&=Y_t[(a_t+\frac{1}{2}b_t^2)\mathrm{d}t+b_t\mathrm{d}B_t]\\&=Y_t(\frac{1}{2}b_t^2\mathrm{d}t+\mathrm{d}X_t).\end{aligned}\qquad(30.16)$$

我们现在来求 $\mathrm{d}Y_t=Y_t\mathrm{d}B_t$ 的伊藤解. 与式(30.16)比较，知 $b_t=1$，$a_t+\frac{1}{2}b_t^2=0$，从而 $a_t=-\frac{1}{2}$. 由式(30.15)，$\mathrm{d}X_t=-\frac{1}{2}\mathrm{d}t+\mathrm{d}B_t$，故 $X_t-X_0=-\frac{t}{2}+B_t$. 由 $Y_t=\mathrm{e}^{X_t}$，知 $Y_t=Y_0\exp(B_t-\frac{t}{2})$. 这与按轨道直接解 $\mathrm{d}Y_t=Y_t\mathrm{d}B_t$ 所得相比，多一因子 $Y_0\mathrm{e}^{-t/2}$.

仿上，式(30.16)的伊藤解可化为先解 $\mathrm{d}X_t=\left(\alpha-\frac{\sigma^2}{2}\right)\mathrm{d}t+\sigma\mathrm{d}B_t$. 它是将式(30.16)与(30.15)比较系数，按式(30.15)得到的. 由定义 30.3.1，$X_t-X_0=\left(\alpha-\frac{\sigma^2}{2}\right)t+\sigma B_t$，故 $L_t=\mathrm{e}^{X_t}$ 有伊藤解

$$L_t/L_0=\exp\left(\left(\alpha-\frac{\sigma^2}{2}\right)t+\sigma B_t\right).\qquad(30.17)$$

(2) 于定理 30.3.6 中令 $n=1$，设 $E\left(\int_\Delta f_t^2 \mathrm{d}t\right)^{1/2} \triangleq E\|f\|_{L^2(\Delta)} < \infty$，则由定理 30.3.6 及 $G_1(t,x)=x$，知 $G_1\left(0+\int_0^t f_s^2 \mathrm{d}s, 0+\int_0^t f_s \mathrm{d}B_s\right)=\int_0^t f_s \mathrm{d}B_s$ 为 $(\mathscr{F}_t)$ 鞅. 由 (30.14)，$E\int_\Delta f_s \mathrm{d}B_s = G_1(0,0)=0$. 这又一次得到性质 30.2.4(4) 的第一个结论.

30.3.2 方程解的存在性与唯一性

先定义条件 H,A,C.

定义 30.3.8 条件 H

(1) $a(t,x)$、$b(t,x)$ 为二元博雷尔可测.

(2) 增长条件(growth conditions)：$\exists$ 常数 k 使 $\forall\, t\in\Delta, x\in R^1$，$a^2(t,x)+b^2(t,x)\leqslant k(1+x^2)$.

(3) 利普希茨(Lipschitz)条件：$\exists$ 常数 L 使

$$|a(t,x)-a(t,y)|+|b(t,x)-b(t,y)|\leqslant L|x-y|,\quad \forall\, t\in\Delta, x,y\in R^1.$$

定义 30.3.9 条件 A

(4) $X(a,\omega)\triangleq A(\omega)$ 不依赖于一切 B_t-B_s，

(5) $EA^2(\omega)<\infty$.

定义 30.3.10 条件 C

(6) 过程 X_t 一切样本在 $t\in\Delta$ 连续.

(7) $\int_a^b EX_t^2 \mathrm{d}t < \infty$.

(8) 对每 $t\in\Delta$，X_t-X_a 为 $\widetilde{\mathscr{F}}_t\triangleq\sigma\{A(\omega), B_s-B_a, a\leqslant s\leqslant t\}$ 可测，即 X_t-X_a 由 t 前的 BM 轨道决定而与将来的值无关.

定理 30.3.11 解的存在与惟一性(existence and uniqueness of solution) 设条件 H 及 A 满足，则方程

$$\begin{cases} \mathrm{d}X_t = a(t,X_t)\mathrm{d}t + b(t,X_t)\mathrm{d}B_t, \\ X(a,\omega)=A(\omega) \quad (\text{初始条件}), \end{cases} \tag{30.18}$$

有满足定义 30.3.10(6),(8)及

$$E(\sup_{t\in\Delta}X_t^2)<\infty \tag{30.19}$$

的随机过程解.且如过程 $\tilde{X}$ 是满足条件 C 的解,则 $\tilde{X}$ 与 X 随机等价:$P(X_t=\tilde{X}_t,t\in\Delta)=1$.

30.3.3 方程的马尔可夫过程解

定理 30.3.12 设 X_t 是定理 30.3.11 确定的过程解,则

(1) 它是马尔可夫过程,初始分布 $P(X_a\in\Gamma),\Gamma\in\mathscr{B}$,转移概率为

$$P(s,x,t,A)=P(X_{s,x}(t)\in A),\quad A\in\mathscr{B}_{[a,b]},$$

其中 $X_{s,x}(t)$ 是下一 SIE 的解

$$\begin{cases}X_t=X_s+\int_s^t a(\tau,X_\tau)\mathrm{d}\tau+\int_s^t b(\tau,X_\tau)\mathrm{d}B_\tau,\\ X_s=x\in R^1\quad[P],\quad a\leqslant s\leqslant t\leqslant b.\end{cases} \tag{30.20}$$

(2) $A_sf(x)=a(s,x)f'(x)+\frac{1}{2}b^2(s,x)f''(x)$,其中 f 是二次连续可微且在 Δ 外为 0 的函数,而算子 A_s 定义如下:

$$A_sf(x)=\lim_{\Delta s\to0}\left[\int_{R^1}f(y)p(s,x,s+\Delta s,\mathrm{d}y)-f(x)\right]/\Delta s.$$

(3) 又如 $a(t,x),b(t,x)$ 与 t 无关,则解 X_t 为齐次马尔可夫过程.

定理 30.3.13 设同定理 30.3.11 且 $a(t,x),b(t,x)$ 为二元连续,$\Delta=[0,b]$,则式(30.18)的解是一扩散过程,$a(t,x),b^2(t,x)$ 分别为其无穷小位移与方差,对任意有界区域中的 t 和 x,一致地有

$$\lim_{h\downarrow0}\frac{1}{h}E(X_{t+h}-X_t|X_t=x)=a(t,x),$$

$$\lim_{h\downarrow0}\frac{1}{h}E((X_{t+h}-X_t)^2|X_t=x)=b^2(t,x),$$

$$E((X_{t+h}-X_t)^4|X_t=x)\leqslant h^2C(t,x),$$

其中 $C(t,x)$ 为此有界区域中一致有界函数.

这样在加于 $a(t,x)$, $b(t,x)$ 相当宽松的条件下,可以构造一个扩散过程.

30.4 多维情形

设 $n\times m$ 维随机矩阵函数 $f(t,\omega)=(f^{ij}(t,\omega))_{n\times m}$ 的元 $f^{ij}(t,\omega)\in M$, $B=(B_t)\triangleq(B_t^1,\cdots,B_t^m)$ 为零初值、m 维 $(\mathscr{F}_t)$BM, $\Delta=[0,T]$. 本节用 f^* 表 f 的转置.

定义 30.4.1 n 维随机向量:

$$I\left(f,\omega\right)\triangleq\int_0^T f_t(t)\omega\Big)\mathrm{d}B_t(\omega)=\left(\sum_{j=1}^m\int_0^T f_t^{1j}\mathrm{d}B_t^j,\cdots,\sum_{j=1}^m\int_0^T f_t^{nj}\mathrm{d}B_t^j\right)$$

称为 f 对 BM 的**随机积分**(SI).

性质 30.4.2 对 m 维 BM 的 SI 的性质

(1) SI 是 f 的线性函数.

(2) 设 $f\in M$ (指 $f^{ij}\in M, i=\overline{1,n}, j=\overline{1,m}$), $E\int_0^T\mathrm{tr}(f(t)f^*(t))\mathrm{d}t<\infty$, 则 $E\int_0^T f(t)\mathrm{d}B_t=0$, $E\left|\int_0^T f(t)\mathrm{d}B_t\right|^2=E\int_0^T\mathrm{tr}(f_t\cdot f_t^*)\mathrm{d}t$.

(3) 如 $f,g\in M$, $E\int_0^T\mathrm{tr}(f_tg_t^*)\mathrm{d}t<\infty$, 则

$$E\left(\int_0^T f_t\mathrm{d}B_t,\int_0^T g_t\mathrm{d}B_t\right)\triangleq E\int_0^T\mathrm{tr}(f_tg_t^*)\mathrm{d}t,$$

这里 $(\cdot,\cdot)$ 表内积.

(4) 对 $\forall f\in M$, 可分向量过程 $\zeta_t=\int_0^t f_s\mathrm{d}B_s$ 以概率 1 样本连续. 如 $x\in R^n$, 令 $\zeta_x(t)=(x,\zeta_t)$, 则对任意正常数 N 和 c^2 有

$$P\left(\sup_{t\in\Delta}|\xi_x(t)|>c\right)\leqslant P\left(\int_0^T(f_tf_t^*x,x)\mathrm{d}t>N\right)+N/c^2.$$

$n=1$ 时

$$P\left(\sup_{t\in\Delta}\left|\int_0^T(f_s,\mathrm{d}B_s)\right|>c\right)\leqslant P\left(\int_0^T|f_t|^2\mathrm{d}t>N\right)+N/c^2.$$

(5) 如 $f\in M, E\int_0^T \mathrm{tr}(f_tf_t^*)\mathrm{d}t<\infty$，则 $\zeta_t=\int_0^t f_s\mathrm{d}B_s, t\in\Delta$ 是 n 维一致平方可积鞅，其特征是 $n\times n$ 阶矩阵过程，且决定于 $\int_0^t f_sf_s^*\mathrm{d}s$. 又对 $\forall\ x\in R^n$，及正数 c，

$$P\left(\sup_{t\in\Delta}|\xi_x(t)|>c\right)\leqslant\frac{1}{c^2}E\int_0^T(f_tf_t^*x,x)\mathrm{d}t, \qquad (30.21)$$

$$E\left(\sup_{t\in\Delta}|\xi_x(t)|^2\right)\leqslant 4E\int_0^T(f_tf_t^*x,x)\mathrm{d}t. \qquad (30.22)$$

特别，$n=1$ 时鞅 $\int_0^t(f_s,\mathrm{d}B_s), t\in\Delta$ 之特征为 $\int_0^t|f_s|^2\mathrm{d}s, t\in\Delta$.

(6) 以 $\mathrm{tr}(f_tf_t^*)$ 易 $|f_t|^2$，定理 30.2.5 成立

设 $n\times m$ 阶矩阵过程 $\sigma_t\in M$，n 维向量过程 X、$\mu\in M_0$ 且 μ 之坐标分量在 Δ 上可积[P]，而对 $0\leqslant s\leqslant t\leqslant T$，有

$$X_t-X_s=\int_s^t\mu_s\mathrm{d}s+\int_s^t\sigma_s\mathrm{d}B_s, \qquad (30.23)$$

则称 n 维过程 X 有**随机微分** $\mathrm{d}X_t=\mu_t\mathrm{d}s+\sigma_t\mathrm{d}B_t$.

公式 30.4.3 多维伊藤变换公式 设 X_t 有 n 维随机微分 $\mathrm{d}X_t$，$\varphi(t,x)$ 为 $\Delta\times R^n$ 上连续实函数，且 $\varphi'_t,\varphi'_{x^i},\varphi''_{x^ix^j}, i,j=\overline{1,n}$ 连续，则 $\varphi(t,X_t)$ 也有 SD，且

$$\mathrm{d}\varphi(t,X_t)=[\varphi'_t(t,X_t)+\mu_t\varphi'_x(t,X_t)+\frac{1}{2}\mathrm{tr}\sigma_t\sigma_t^*\varphi''_{xx}(t,X_t)]\mathrm{d}t$$
$$+(\varphi'_x(t,X_t),\sigma_t\mathrm{d}B_t). \qquad (30.24)$$

这里 $\qquad \varphi'_x\triangleq(\varphi'_{x^1},\varphi'_{x^2},\cdots,\varphi'_{x^n}), \varphi''_{xx}=(\varphi''_{x^ix^j})_{n\times n}$.

定理 30.4.4 多维 SIE 解的存在与惟一性 设 $a(t,x)$ 和 $b(t,x)$ 为 $\Delta\times R^m$ 上分别取值于 R^m 和 $L(R^m)\triangleq\{R^m$ 到 R^m 的一切线性算子$\}$的可测函数，且

$$|a(t,x)|^2 + |b(t,x)|^2 \leqslant K(1 + |x|^2), x \in R^m, \tag{30.25}$$

$$|a(t,x) - a(t,y)|^2 + |b(t,x) - b(t,y)|^2 \leqslant C_R|x - y|^2,$$
$$|x| \leqslant R, |y| \leqslant R, \tag{30.26}$$

这里 K, C_R 为某常数，R 预先任意给定的正数. $|x|^2 = \sum^m (x^i)^2, x = (x^1, \cdots, x^m), |b(t,x)|^2 = \sum_{i,j=1}^m |b^{ij}(t,x)|^2$. 又设 $X_0(\omega) = A(\omega) \in \mathscr{F}_0$，则 SDE

$$\begin{cases} \mathrm{d}X_t = a(t,X_t)\mathrm{d}t + b(t,X_t)\mathrm{d}B_t, \\ X_0 = A, \end{cases} \tag{30.27}$$

有惟一(随机等价意义下)连续解 $X_t = (X_t^1, \cdots, X_t^m)$. 且为马尔可夫过程. 这里"解"指 X 为 $(\mathscr{F}_t)_{t\in T}$ 适应，$\int_0^T |X_t^i|\mathrm{d}t < \infty \quad [P]$，$\sigma(t,X_t)$ 的元 $\sigma^{ij}(t,X_t)$ 在 $[0,T]$ 平方可积 $[P]$，且 X_t 有 SD(30.27) 和 $X_0 = A$.

定理 30.4.5 设同定理 30.4.4 且 $a(t,x), b(t,x)$ 为连续，则 (30.27) 的解是一扩散过程，漂移系数向量 $\mu(t,x) = a(t,x)$，扩散矩阵 $\sigma^2(t,x) = b(t,x)b^*(t,x)$. 初始分布为 $P(A \in \Gamma)$，而转移概率 $P(s,x;t,\Gamma) = P(X_{s,x}(t) \in \Gamma)$，$X_{s,x}(t)$ 是

$$\xi(t) = x + \int_s^t a(\tau, \xi(\tau))\mathrm{d}\tau + \int_s^t b(\tau, \xi(\tau))\mathrm{d}B_\tau, 0 \leqslant s < t \leqslant T$$

的解.

定理 30.4.6 设同定理 30.4.4 且 $a(t,x), b(t,x)$ 连续，且对 x 二次连续可微. 又对某 $p > 0$ 及 $k > 0$,

$$\sum_{i,k=1}^m \left|\frac{\partial a^i(t,x)}{\partial x^k}\right| + \sum_{i,j,k=1}^m \left|\frac{\partial^2 a^i(t,x)}{\partial x^j \partial x^k}\right| + \sum_{i,j,k=1}^m \left|\frac{\partial b^{ij}(t,x)}{\partial x^k}\right|$$
$$+ \sum_{i,j,k,l=1}^m \left|\frac{\partial^2 b^{ij}(t,x)}{\partial x^k \partial x^l}\right| \leqslant k(1 + |x|^p),$$

而 $\varphi(x)$ 为 R^m 上二次连续可微实函数，满足

$$|\varphi(x)| + \sum_{i=1}^{m}\left|\frac{\partial\varphi(x)}{\partial x^i}\right| + \sum_{i,j=1}^{m}\left|\frac{\partial^2\varphi(x)}{\partial x^i\partial x^j}\right| \leqslant k(1+|x|^p).$$

则函数 $u(s,x)\triangleq E\varphi(X_{s,x}(t))$，$0\leqslant s\leqslant t\leqslant T$，$x\in R^m$ 满足科尔莫戈罗夫向后方程，这里 $X_{s,x}(t)$同(30.4.5)

$$\frac{\partial u(s,x)}{\partial s} + \sum_{i=1}^{m}a^i(s,x)\frac{\partial u(s,x)}{\partial x^i} + \frac{1}{2}\sum_{i,k=1}^{m}\sum_{j=1}^{m}b^{ij}(s,x)b^{kj}(s,x)\frac{\partial^2 u(s,x)}{\partial x^i\partial x^j} = 0.$$

及边界条件 $\lim_{s\uparrow t}u(s,x) = \varphi(x)$.

30.5 随机微分方程的模型与应用

30.5.1 白噪声

BM 的协方差函数 $\Gamma_B(s,t)=\min(s,t)$是连续的，因此(B_t)是均方连续的. 但偏导数$\frac{\partial^2}{\partial s\partial t}\Gamma_B(s,t)=\frac{\partial}{\partial s}I_{(t,\infty)}(s)$不存在，故$(B_t)$均方导数不存在. 但形式上有 $\Gamma_{\dot B}(s,t)=\frac{\partial^2}{\partial s\partial t}\Gamma(s,t)=\delta(s-t)$，它形式上表示 $\dot B(s)$与 $\dot B(t)$在 $s\neq t$ 时不相关，这里 $\delta(\cdot)$为狄拉克(Dirac)函数. 称$\frac{\mathrm{d}B_t}{\mathrm{d}t}\triangleq W_t$，$t\in R_+\geqslant 0$，为(正态)**白噪声过程**(white noise process). 它不是一个实际的物理过程，而是一种数学的抽象：它的平均功率是无限的，功率谱密度是一个非零常数.

注意 B 是高斯过程. 如 Y 是一个高斯过程，其均值为 0，协方差函数 $\Gamma_Y(s,t)=EY_sY_t$ 有连续二阶偏导 $\rho(s,t)=\frac{\partial^2}{\partial s\partial t}\Gamma_Y(s,t)$，则其均方意义下导函数 $\dot Y_t$ 对所有 t 存在，且仍为正态，$\Gamma_{\dot Y}(s,t)=E\dot Y_s\dot Y_t=\rho(s,t)$. 故适当选择 Y，则 $\dot Y$ 可作为白噪声的一种近似. 白噪声 $\dot B_t=W_t$ 可视为连续时间平稳高斯过程，均值为 0，协方差函

数为 $\delta(s-t)$.

在实用上如某随机信号，在实际系统的通频带范围内具有较平坦的功率谱密度，就可近似作白噪声处理.

30.5.2 随机动力系统

设 f 为一光滑的(即 f' 连续)实值函数，W_t 为一随机分布函数，称下一形式的 SDE 的解为一连续的随机动力系统(continuous stochastic dynamical system)：

$$\frac{\mathrm{d}X_t}{\mathrm{d}t}=f(t,X_t,W_t), \tag{30.28}$$

初值条件 X_0 固定或是一确知其分布的随机变量. W_t 也叫做随机输入因子. 在涉及随机控制、滤波、外推及预报研究的物理工程模型，会经常遇到这类问题(见第 31 章).

只涉及一个非齐次随机成分时，(30.28)有如下形式

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t=f(t,X_t)+W_t. \tag{30.29}$$

由白噪声导出的线性振荡器，其移动位置对应一个二阶 SDE $\left(以\ \dot{X}_t\ 记\frac{\mathrm{d}}{\mathrm{d}t}X_t\right)$

$$\ddot{X}_t+2\beta\dot{X}_t+\gamma^2X_t=\dot{B}_t\triangleq W_t, \tag{30.30}$$

这里阻尼参数 β 与惯性常数 γ 均为实的常数. 若令 $X_2(t)=\dot{X}_1(t)$，$\boldsymbol{X}(t)=(X_1(t),X_2(t))^{\mathrm{T}}$，$\boldsymbol{W}_t=(0,W_t)^{\mathrm{T}}$，则(30.30)可写为向量形式：

$$\dot{\boldsymbol{X}}_t=\begin{pmatrix}0 & 1\\ -\gamma^2 & -2\beta\end{pmatrix}\boldsymbol{X}_t+\boldsymbol{W}_t, \tag{30.31}$$

从而可视为式(30.29)的一个特例.

1. 奥恩斯坦-乌伦贝克过程

设 Y_t 为时刻 t 在液面上运动的一粒子的某方向上的速度，m

为其质量，$-fY_t$ 表示粒子与液体摩擦而产生的阻尼力，则由牛顿第二定律，可知

$$m\dot{Y}_t = -fY_t + W_t, \quad t \geqslant 0, f > 0. \tag{30.32}$$

这里 W_t 是分子冲击粒子所产生的总合力，设为正态白噪声. 令 $\beta = f/m$，又可写为

$$\mathrm{d}Y_t = -\beta Y_t \mathrm{d}t + \mathrm{d}B_t, \quad \beta > 0, t > 0 \text{ 而 } Y_0 = 0. \tag{30.33}$$

由定理 30.3.13，其解是一扩散过程，漂移系数 $a(t,y) = -\beta y$，扩散系数 $b^2(t,y) = 1$，直接解得

$$Y_t = \int_0^t \mathrm{e}^{-\beta(t-s)} \mathrm{d}B_s \tag{30.34}$$

由节 25.3.4 知为奥恩斯坦-乌伦贝克(Ornstein-Uhlenbeck)过程，其转移密度为

$$p(t,x,y) = \varphi\left(\frac{1}{2\beta}(1-\mathrm{e}^{-2\beta t}), \quad x\mathrm{e}^{-\beta t}, y\right), \tag{30.35}$$

其中 $\varphi(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp(-(y-x)^2/2t), t>0, x,y \in R^1$.

若以 X_t 表示 t 时质点位置，则由 $X_t - X_0 = \int_0^t Y_s \mathrm{d}s$ 及定理 22.4.6 知，X_t 为正态，$X_t \sim N(0, (\beta t - 1 + \mathrm{e}^{-\beta t})/\beta^3)$.

2. 简单信号防护模型

电力动态系统常遇如下 SDE：

$$\mathrm{d}X_t = F_t X_t \mathrm{d}t + G_t \mathrm{d}B_t \quad (\text{线性})$$

和

$$\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}B_t \quad (\text{非线性})$$

可观察的过程受控于

$$\mathrm{d}Z_t = H_t X_t \mathrm{d}t + R_t \mathrm{d}\beta_t.$$

这里 B 与 β 为独立的 BM，详见[91]、[92].

30.5.3 随机环境中的群体增长

大多数生态模型可写成

$$\dot{N}_t = f(t, N_t) + g(t, N_t)W_t, \quad W_t = \dot{B}_t, \tag{30.36}$$

其中 N_t 为 t 时群体的大小，最简单的是“指数”增长方程

$$\dot{N}_t = (\alpha + \gamma W_t)N_t, \text{或 } \mathrm{d}N_t = \alpha N_t \mathrm{d}t + \gamma N_t \mathrm{d}B_t \tag{30.37}$$

α, γ 为常数，$\alpha + \gamma W_t$ 为瞬时增长率，这里考虑到了随机环境的影响. 此方程在伊藤积分意义下的解是 $(0, \infty)$ 上扩散过程，其漂移与扩散系数分别为 $\mu(x) = \alpha x$ 和 $\sigma^2(x) = \gamma^2 x^2$. 由(30.16)也可得此解为一拓广的几何 BM，$N_t = \exp\left(\left(\alpha - \frac{1}{2}\gamma^2\right)t + \gamma B_t\right)$.

若改写(30.37)为 $\mathrm{d}N_t/N_t = \alpha \mathrm{d}t + \gamma \mathrm{d}B_t$，则两边积分解得 $N_t = N_0 \exp(\alpha t + \gamma B_t)$，这是 $\mu(x) = \left(\alpha + \frac{1}{2}\gamma^2\right)x$ 和 $\sigma^2(x) = \gamma^2 x^2$ 的另一个几何 BM. 漂移系数多出 $\frac{1}{2}\gamma^2 x$ 项是白噪声造成的. 在 S 积分意义下从(30.36)可准确算出这些系数. 在 $t \to \infty$ 时，如 $0 < \alpha < \frac{1}{2}\gamma^2$，则伊藤意义下的解过程 $N_t \to 0$，而后者 S 积分意义，$N_t \to \infty$.

例 30.5.1 随机逻辑斯谛方程 一种生物聚居地的承受能力可描述为

$$\dot{N}_t = N_t(\alpha + \gamma \dot{B}_t - N_t),$$

其伊藤解是 $\mu(x) = x(\alpha - x)$ 和 $\sigma^2(x) = \gamma^2 x^2$ 的 $(0, \infty)$ 上的扩散过程，而 S 意义下的解的漂移系数也增加了 $\frac{x}{2}\gamma^2$ 一项. 后者恒有平稳概率分布密度，$EN_\infty = \alpha$，$V(N_\infty) = \frac{1}{2}\alpha\gamma^2$，而前者只在 $2\alpha/\gamma^2 > 1$ 时才有平稳分布.

30.5.4 基因频数波动

有两类生物 A_1 和 A_2 组成的一个大群体；在 t 时 A_1 的比重为 $p(t)$，A_1 对 A_2 的固有遗传选择优势为 $1+s$ 比 1，则基因频数变化的确定性方程为 $\dot{p}(t) = sp(1-p)$. 将选择微分 $s\mathrm{d}t$ 以白噪声微分

$d\widetilde{B}_t$ 代替，并要求 $E(d\widetilde{B}_t)=a dt$，$E(d\widetilde{B}_t)^2=\nu^2 dt$，则

$$dp_t = p_t(1-p_t)d\widetilde{B}_t, \tag{30.38}$$

有一扩散过程解，系数 $\mu(x)=ax(1-x)$，$\sigma^2(x)=\nu^2x^2(1-x^2)$.

30.5.5 动态经济问题

最近20年间，用SDE描述和解决动态经济问题有了很大发展. 传统的经济数学模型致力于生产和消费因素间瞬时的和平衡的相互关系，而SDE提供一条途径，把作用于多经济单元（劳动力，技术变量，股票价格等等）的随机因素、不定因素及危险因素，都协调起来考虑. 其次，用连续SDE来刻画随机波动，常可作更为精辟的分析，容易阻止和预防那些敏感弊端出现.

主要用于三类现象：(1)在不确定条件下描述确定因素的增长；(2)在市场条件不变情况下优化价格变量的本质；(3)随机动态规划与控制目标.

1. 单一经济的部门中不确定条件下的生产与消费变量

令 K_t 为 t 时资本资产，L_t 为可用劳动力，C_t 为消费率. 在许多场合要求生产函数 $F(K,L)$ 对 K,L 是凸的和齐次的，熟知的科布-道格拉斯(Cobb-Douglas)生产函数 $F(K,L)=K^{\alpha}L^{1-\alpha}$，$0<\alpha<1$ 即一例. 动产累积方程为

$$\dot{K}_t = F(K_t,L_t) - \lambda K_t - C_t, \tag{30.39}$$

此处 λ 为资金的贬值率. 由于人口数量从而 L_t 是受随机干扰支配的，故设

$$dL_t = L_t[\alpha dt + \sigma dB_t], \tag{30.40}$$

式(30.17)已给出解

$$L_t/L_0 = \exp((\alpha-\sigma^2/2)t+\sigma B_t),$$

这是有指数漂移的几何BM.

2. 优化累积模型

生产过程设为 $X_t=F(K_t,L_t,Z_t)$，Z_t 为 t 时技术，并设 δ,γ 为常数及

$$\begin{cases}\dot{K}_t = X_t - C_t - \delta K_t (\text{累积方程}), \\ \dot{L}_t = \gamma L_t, \\ \mathrm{d}Z_t = \pi(Z_t)\mathrm{d}t + \sigma(Z_t)\mathrm{d}B_t. \end{cases} \tag{30.41}$$

典型的随机控制问题是找符合上述条件的 C_t，使 $E\int_0^\infty U(C_t)\mathrm{e}^{-\rho t}\mathrm{d}t$ 最大，这里 $U(C_t)$ 是消费率为 C_t 时的对个人所应设置的公用事业设备费.

30.6 对鞅的随机积分方程

存在定理 30.4.4 中利普希茨(Lipschitz)条件太强了，许多问题中不能满足. 另一方面，早在 50 年代初杜布就注意到可仿对 BM 的伊藤积分给出对某些鞅的随机积分，这只要在条件 30.6.1 成立的假定下，以 $(\mathscr{F}_t)$ 鞅 M_t 和 $\int_0^t f_s\mathrm{d}s$ 分别易 $(\mathscr{F}_t)$BM B_t 和 t，就可实现.

条件30.6.1 $M=(M_t)$ 为 $(\mathscr{F}_t)$ 鞅，且 $\exists$ $(\mathscr{F}_t)$ 适应过程 (f_t)，使 $E[(M_t-M_s)^2|\mathscr{F}_s]=E\left(\int_s^t f_u\mathrm{d}u|\mathscr{F}_s\right)$.

略作推广，以非负 $(\mathscr{F}_t)$ 适应过程 (F_t) 代替 $\left(\int_0^t f_u\mathrm{d}u\right)$，并要求 M_t，$M_t^2-F_t$ 都是 $(\mathscr{F}_t)$ 鞅就够了. 60年代迈耶(Meyer)证明了对平方可积鞅，上述过程 (F_t) 存在，从而向现代鞅论及随机积分迈出关键一步.

条件30.6.2 设 $t\in\Delta=[0,T]$，$x\in R^m$，

(1) 取值于 $L(R^m)$(见定理30.4.4)的 $b(t,x)$ 的所有元素 $b^{ij}(t,x)$ 是 (t,x) 的连续函数，且满足 $(b(t,x)\theta,\theta)\leqslant c|\theta|^2$，

$x,\theta\in R^m$，c 为某正常数.

(2) 对$\forall$ $(t,x)\in\Delta\times R^m$，矩阵 $b(t,x)=(b^{ij}(t,x))$正定，即对一切非零 θ，$(b(t,x)\theta,\theta)>0$.

(3) $m+1$元函数 $a(t,x)$可测且有界.

定理30.6.3 设 Ω 为 $\Delta=[0,T]$到 R^m 的连续函数全体，$\mathscr{N}_T^s=\sigma\{X_t,s\leqslant t\leqslant T\}$，$X_t\in\Omega$. 又设条件30.6.2成立，则对$\forall$ $s\in\Delta$ 及 $x\in R^m$，在$(\Omega,\mathscr{N}_T^s)$上存在惟一概率测度 $P_{sx}^{a,b}$使

$$P_{sx}^{a,b}(X_s=x)=1. \tag{30.42}$$

过程 $X_t-X_s-\int_s^t a(u,X_u)\mathrm{d}u,t\in[s,T]$，是关于$(\mathscr{N}_T^s,P_{sx}^{a,b})$为平方可积鞅，有马尔可夫性，且以$\int_a^b(u,X_u)\mathrm{d}u$ 为其特征.

设 $Y=(Y_t)$为$[s,T]$上 $n\times m$ 维过程，$(\mathscr{F}_t)$适应，且

$$P_{sx}^{a,b}\left(\int_S^T\mathrm{tr}(Y_tb(t,X_t)Y_t^*)\mathrm{d}t<\infty\right)=1,$$

则仿对BM的积分定义 30.4.1 及 30.2.3，可定义 Y 对平方可积鞅 $M_t\triangleq X_t-X_s-\int_s^t a(u,X_u)\mathrm{d}u$ 的随机积分为

$$I(Y,M)=\int_s^T Y_t\mathrm{d}M_t.$$

如 $E_{sx}^{a,b}\int_S^T\mathrm{tr}[Y_tb(t,X_t)Y_t^*]\mathrm{d}t<\infty$，则可分过程 $\zeta_t(M)\triangleq\int_s^t\overline{Y}_\tau\mathrm{d}M_\tau,t\in[s,T]$，关于$(\mathscr{N}_t^s,P_{sx}^{a,b})$ 是连续平方可积鞅，其特征为

$$\langle\zeta(M)\rangle_t=\int_s^t Y_\tau b(\tau,X_\tau)Y_\tau^*\mathrm{d}\tau.$$

特别令 $Y_t=b^{\frac{1}{2}}(t,X_t)$，则过程

$$B_t\triangleq\int_s^t b^{-\frac{1}{2}}(\tau,X_\tau)\mathrm{d}M_\tau,t\in[s,T],$$

关于$(\mathscr{N}_t^s,P_{sx}^{a,b})$是BM，且

$$M_t=\int_s^t b^{\frac{1}{2}}(\tau,X_\tau)\mathrm{d}B_\tau,t\geqslant s. \tag{30.43}$$

这样在30.6.2条件下，对$\forall\ s\in[0,T]$，$x\in R^m$，在Ω上存在m维BM$(B_t)_{t\geq s}$，使

$$X_t = X_s + \int_s^t a(\tau,X_\tau)\mathrm{d}\tau + \int_s^t b^{\frac{1}{2}}(\tau,X_\tau)\mathrm{d}B_\tau.$$

从而在$a(t,x)$、$b(t,x)$足够光滑条件下，可构造SDE的扩散过程解(常称为弱解).

31 预报与滤波

31.1 预报理论

若 $X_T=\{X(t),t\in T\}$(T 为整数集或全体实数)为平稳过程(23.1),当随机变量 $X(t)$直到时刻 t 为止的值 $X(s)(s\leqslant t)$已被观测时,要研究的问题是以这些观测值为基础,对 $t+\tau(\tau>0)$时间的值进行预报. **预报理论**(prediction theory)是由科尔莫戈罗夫与维纳所开创,后来又有许多人研究. 这个理论是平稳过程理论当中的一个重要分支,它应用于通信、电工学、气象、水文以及经济问题中.

设对 $X(t+\tau)$的预报为 $\hat{X}_t(\tau)$,预报误差可以选用不同形式:均方误差、条件概率及误差概率等. 在工程技术中广为采用的是均方误差. 这样,预报标准就是要使 $E|X(t+\tau)-\hat{X}_t(\tau)|^2$ 最小. 显然 $\hat{X}_t(\tau)$应是序列 $X(s)$, $s\leqslant t$ 的某个函数(叫预报函数 prediction function)$g(\tau,t)\triangleq g_\tau(X_s,s\leqslant t)$. 由于高斯序列是实用中最大量、最有代表性的,而在 X_s, $s\leqslant t$ 的所有线性函数中,如某线性函数 $g(t,\tau)$使预报的均方误差最小,则对所有 X_s, $s\leqslant t$ 的函数用作预报,$g(t,\tau)$也达最小. 因此本章主要考虑采用均方误差的线性预报.

与线性预报理论在问题的提出与解决方法上都有许多类似之处的,还有插值(interpolation)(本章略)与滤波(filtering). 滤波产生于从混有噪声的信号中分离出所需要的成分的通信问题.

定义 31.1.1 平稳线性最优预报. 设$\{X(t),t\in T\}$为概率空间$(\Omega,\mathscr{F},P)$上的弱平稳过程,且$E(X(t))\equiv 0$. 令 $\mathscr{H}_t$ 为使$\{X(s)|$

$s \leqslant t\}$可测的最小σ代数，$H(t)$为$\mathscr{N}_t$可测且关于概率测度P平方可积的函数全体所构成的希尔伯特空间，$M_t(X)$为由$\{X(s), s \leqslant t\}$所组成的线性子空间. 且

$$\bigcup_t H_t(X) = H(X), \quad \bigcup_t M_t(X) = M(X).$$

当$\{X(s), s \leqslant t\}$已知时，由$H_t(X)$或$M_t(X)$的元$\hat{X}_t(\tau)$对$X(t+\tau)$ $(\tau > 0)$进行预报，这就是预报理论问题，称$\hat{X}_t(\tau)$为$X(t+\tau)$的**预报值**(prediction).

定义 31.1.2 在预报值中使预报均方误差

$$\sigma^2(\tau) = E(|X(t+\tau) - \hat{X}_t(\tau)|^2)$$

最小的$\hat{X}_t(\tau) \in H_t(X)$，称为**均方最优预报**(mean square optimum prediction). 当所求的$\hat{X}_t(\tau)$限于$M_t(X)$时，称为**最优线性预报**(optimum linear prediction)或者**外推**(extrapolation).

由$\{X(t), t \in T\}$的平稳性，$\sigma^2(\tau)$与t无关.

线性预报的最优预报值是$X(t+\tau)$向$M_t(X)$的投影.

若T为整数集，$X(t)$有谱分解

$$X_t \triangleq X(t) = \int_{-\pi}^{\pi} e^{i\lambda t} d\zeta(\lambda).$$

这时必存在关于谱测度$dF(\lambda) = E(|d\zeta(\lambda)|^2)$平方可积的函数$\hat{\varphi}(\lambda, \tau)$，能将$\hat{X}_t(\tau)$表示为

$$\hat{X}_t(\tau) = \int_{-\pi}^{\pi} e^{i\lambda t} \hat{\varphi}(\lambda, \tau) d\zeta(\lambda). \qquad (31.1)$$

一般，$M_s(X)$与s同时增加，而当$M_s(X)$与s无关时，称X_t为**确定性**的(deterministic). 当$\bigcap_s M_s(t) = \{0\}$时，称为**非确定性**的，此时则可求得$\hat{\varphi}(\lambda, \tau)$，使对一切的$t, \tau$，有$\hat{X}_t(\tau) = X(t+\tau)$.

如果X_t是纯**非确定性**的，则其谱测度关于勒贝格(Lebesgue)测度绝对连续，其密度函数$f(\lambda)$几乎处处为正，并且

$$\int_{-\pi}^{\pi} \log f(\lambda) d\lambda > -\infty, \qquad (31.2)$$

且反之也真.

下面用此 $f(\lambda)$来给出最优预报值，存在属于关于单位圆的哈代(Hardy)族 H_2 的

$$r(z)=\sum_{t=0}^{\infty}C(t)z^t, \tag{31.3}$$

用其边界函数 $\varphi(\lambda)=r(\mathrm{e}^{-i\lambda})$将 $f(\lambda)$表示为

$$f(\lambda)=\frac{1}{2\pi}|\varphi(\lambda)|^2. \tag{31.4}$$

而 X_t 可用白噪声序列$\{\varepsilon_t,t\in T\}$和 $r(z)$的系数 $C(t)$表示为

$$X_t=\sum_{s=-\infty}^{t}C(t-s)\varepsilon_s. \tag{31.5}$$

这是后向滑动和模型. 这里满足(31.5)的$\{C(t)\}$和$\{\varepsilon_t\}$的组虽有多种，但当 $r(z)$为极大时，

$$r(z)=\sqrt{2\pi}\exp\left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\log+(\lambda)\frac{\mathrm{e}^{-i\lambda}+z}{\mathrm{e}^{-i\lambda}-z}\mathrm{d}\lambda\right). \tag{31.6}$$

(31.5)的表示在下述意义上是标准的：即对任意 t 有 $M_t(X)=M_t(\varepsilon)$. 因而 $\hat{X}_t(\tau)$为 $X(t+\tau)$向 $M_t(\varepsilon)$的投影，

$$\hat{X}_t(\tau)=\sum_{s=-\infty}^{\infty}C(t+\tau-s)\varepsilon_s \tag{31.7}$$

$$\left(=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda t}\hat{\varphi}(\lambda,\tau)\mathrm{d}\zeta(\lambda)\right).$$

并且 $$\hat{\varphi}(\lambda,\tau)=\mathrm{e}^{i\lambda\tau}\left(\varphi(\lambda)-\sum_{s=0}^{\tau-1}C(s)\mathrm{e}^{-i\lambda s}\right)\Big/\varphi(\lambda),$$

预报误差为

$$\sigma^2(\tau)=\sum_{s=0}^{\tau-1}|C(s)|^2. \tag{31.8}$$

31.2 各类时序模型的预报方法

31.2.1 AR(p)序列的预报方法

设 AR(p)序列为 $\varphi(B)X_t=\varepsilon_t$，其中

公式 31.2.1 $\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p$，$\varepsilon_t$ 为白噪声序列. 预报方法如下

1. 递推公式法

关于 l 的递推预报公式为

公式 31.2.2 $\begin{cases}\varphi(B)\hat{X}_t(l)=0, & l>0,\\ \hat{X}_k(-l)=X_{k-l}, & l\geqslant 0.\end{cases}$

后一式给出初值. l 越大时，预报准确性越差. 这里 φ_i 认为已知.

2. 预报函数法

将 $\hat{X}_k(l)$ 可写为一般形式

$$\hat{X}_k(l)=\sum_{i=1}^{p} b_i^{(k)} f_i(l),\quad l>-p. \tag{31.9}$$

初值为 $\hat{X}_k(-j)=X_{k-j}$，$j=0,1,\cdots,p-1$，即

$$\sum_{i=1}^{p} b_i^{(k)} f_i(-j)=X_{k-j},\quad j=0,1,\cdots,p-1.$$

由此求出 $b_i^{(k)}$，代入(31.9)所得，即为预报函数. 由此可方便地作出预报.

由公式 31.2.2 和式(31.9)可知，对于 AR(p)序列，其预报值 $\hat{X}_k(l)(l\geqslant 0)$ 只需用到 $X_k,X_{k-1},\cdots,X_{k-p+1}$ 的值. 而用不着更早的历史值 $X_j(j\leqslant k-p)$.

例 31.2.3 求 $(1-\varphi_1 B-\varphi_2 B^2)X_t=\varepsilon_t$ 的预报.

解 由公式 31.2.2 可得

$$\hat{X}_k(l)=\varphi_1\hat{X}_k(l-1)+\varphi_2\hat{X}_k(l-2),\quad l>0,$$

初值 $\hat{X}_k(0)=X_k$，$\hat{X}_k(-1)=X_{k-1}$.

下面找预报函数. 令

$$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2 \triangleq (1-z_1 B)(1-z_2 B),$$

即 z_1,z_2 为 $\varphi(B)=0$ 的两个根，

当 z_1,z_2 为两个不同实根时，由例 23.7.6

$$\hat{X}_k(l) = b_1^{(k)} z_1^l + b_2^{(k)} z_2^l, \quad l > -2$$

其中 $b_1^{(k)}, b_2^{(k)}$ 满足方程

$$\begin{cases} b_1^{(k)} + b_2^{(k)} = X_k & (\text{取 } l = 0), \\ b_1^{(k)} z_1^{-1} + b_2^{(k)} z_2^{-1} = X_{k-1} & (\text{取 } l = -1). \end{cases}$$

由此解出

$$b_1^{(k)} = \frac{X_k - z_2 X_{k-1}}{1 - z_2 z_1^{-1}}, \quad b_2^{(k)} = \frac{X_k - z_1 X_{k-1}}{1 - z_1 z_2^{-1}},$$

得预报函数

$$\hat{X}_k(l) = \frac{z_1^{l+1} - z_2^{l+1}}{z_1 - z_2} X_k - \frac{z_1 z_2 (z_1^l - z_2^l)}{z_1 - z_2} X_{k-1}.$$

若 $z_1 = z_2$，由例 23.7.6 可得

$$\hat{X}_k(l) = (l+1) X_k z_1^l - l X_{k-1} z_1^{l+1}.$$

若 z_1, z_2 为一对共轭复根，$z_1 = \rho(\cos\alpha + i\sin\alpha)$，$z_2 = \bar{z}_1$，可得

$$\hat{X}_k(l) = b_1^{(k)} \rho^l \cos l\alpha + b_2^{(k)} \rho^l \sin l\beta.$$

由初值解出 $b_i^{(k)}$，$i = 1,2$，预报函数为

$$\hat{X}_k(l) = X_k \rho^l \frac{\sin(l+1)\alpha}{\sin\alpha} - X_{k-1} \rho^{l+1} \frac{\sin l\alpha}{\sin\alpha}.$$

31.2.2 MA(q)序列的预报方法

设 X_t 为 MA(q)序列，满足 $X_t = \theta(B)\varepsilon_t$，其中

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q. \qquad (31.10)$$

预报方法如下：

1. 直接方法

MA(q)序列的逆转形式为 $X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + \varepsilon_t$，其中 I_j 是下一展式的系数

$$\theta^{-1}(B) = 1 - \sum_{j=1}^{\infty} I_j B^j.$$

预报公式为

$$\hat{X}_k(l) = \begin{cases} \sum_{j=1}^{\infty} I_j^{(l)} X_{k+1-j}, & 0 \leqslant l \leqslant q, \\ 0, & l > q. \end{cases} \tag{31.11}$$

其中

$$\begin{cases} I_j^{(l)} = \sum_{i=1}^{l-1} I_j^{(l-i)} I_i + I_{j+l-1}, & 0 \leqslant l \leqslant q, \\ I_j^{(q)} = I_j, & l > q. \end{cases}$$

由公式(31.10)可知，$\hat{X}_k(l)$要用到全部历史数据 $X_{k-j}(j \leqslant 0)$. 但是真实数据是有限的，实际应用时只是有穷和的近似，即

$$\hat{X}_k(l) \approx \sum_{j=1}^{M} I_j^{(l)} X_{k+1-j}, \quad 0 \leqslant l \leqslant q. \tag{31.12}$$

其中 M 为采样数. 如果 MA 序列是可逆的，I_j 便具有负指数的收敛性，因而 $I_j^{(l)}$ 也是负指数衰减的. 由 I_j 的收敛速度适当地选取 M，以保证其预报精度.

2. 预报矢量的递推算法

定义 31.2.4 称

$$\hat{X}_k^q = (\hat{X}_k(1), \hat{X}_k(2), \cdots, \hat{X}_k(q))^{\mathrm{T}}$$

为 MA 序列的**预报矢量**(prediction vector)

$\hat{X}_k^q$ 和 $\hat{X}_k(l)=0(l>q)$一起描述了在 k 时刻对未来的全部预报结果. 由于 MA 序列 $G_l=-\theta_l$,$(1 \leqslant l \leqslant q)$，其中 G_l 为格林函数，应用

$$\hat{X}_{k+1}(l) = \hat{X}_k(l) + G_l(X_{k+1} - \hat{X}_k(1)),$$

得到 MA 序列的递推公式为

$$\hat{\boldsymbol{X}}_{k+1}^q = \begin{pmatrix} \theta_1 & 1 & 0 & \cdots & 0 \\ \theta_2 & 0 & 1 & & \\ \vdots & & & \ddots & \\ \theta_q & 0 & 0 & \cdots & 1 \end{pmatrix} \hat{\boldsymbol{X}}_k^q - \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{pmatrix} \boldsymbol{X}_{k+1}. \tag{31.13}$$

递推的初值可取某时刻 k_0 的 $\hat{X}_{k_0}^q$，它可以直接求出. 当 k_0 很

小时，也可取为 0，由于(31.13)的迭代格式是稳定的，当 k 增大到很大以后，初值的误差影响将会逐渐消失.

31.2.3 ARMA(p,q)序列的预报方法

设 X_t 为平稳可逆 ARMA(p,q)序列(23.7.1)，可写 $\varphi(B)X_t=\theta(B)\varepsilon_t$，其中 $\varphi(B)$，$\theta(B)$分别由(31.2.1)及(31.10)定义.

1. 直接方法

ARMA(p,q)序列的直接预报方法就是从模型本身出发进行预报.有三种等价形式：

(1) 由差分方程预报

公式 31.2.5

$$\hat{X}_k(l)-\sum_{j=1}^{l-1}\varphi_j\hat{X}_k(l-j)=\sum_{j=1}^{p}\varphi_jX_{k-j}+\sum_{j=1}^{q}\theta_j\varepsilon_{k-j+1},$$

上式中若求和上限小于下限，约定此和为 0，下同.公式(31.13)是以 $\hat{X}_k(1),\hat{X}_k(2),\cdots,\hat{X}_k(p),\cdots,\hat{X}_k(l)$ 为变元的线性方程组，它的系数矩阵是下三角阵，因此只需迭代求解，这在计算机上容易实现.

(2) 格林函数预报法

预报公式为

公式 31.2.6

$$\hat{X}_k(l)=\sum_{j=0}^{\infty}G_{l+j}\varepsilon_{k-j},$$

其中格林函数 G_l 可由 ARMA 模型的参数递推算出.

(3) 逆函数预报法

预报公式为

公式 31.2.7

$$\hat{X}_k(l)=\sum_{j=1}^{\infty}\Pi_j^{(l)}X_{k+1-j}=\Pi^{(l)}(B)X_k,$$

其中 $\Pi^{(l)}(B)=\sum_{j=1}^{\infty}\Pi_j^{(l)}B^{j-1}$，而 $\Pi_j^{(l)}$ 由下式给出

$$\begin{cases}\Pi_j^{(l)}=\Pi_{j+l-1}+\sum_{i=1}^{l-1}\Pi_i\Pi_j^{(l-i)},l>1,\\ \Pi_j^{(1)}=\Pi_j,\end{cases}$$

这里 Π_j 由 $\varepsilon_t=\sum_{j=0}^{\infty}\Pi_jB^jX_{t-j}$ 决定的系数，$\Pi_0=1$.

2. 预报矢量的递推算法

ARMA(p,q)序列的预报矢量为

$$\hat{\boldsymbol{X}}_k^q=(\hat{X}_k(1),\hat{X}_k(2),\cdots,\hat{X}_k(q))^{\mathrm{T}},$$

预报公式为

公式 31.2.8

$$\hat{X}_{k+1}(l)=\boldsymbol{G}\hat{X}_k(l)+\boldsymbol{\alpha}X_{k+1}+\boldsymbol{\beta},$$

其中

$$\boldsymbol{G}=\begin{bmatrix}-\psi_1 & 1 & 0 & \cdots & 0 & 0\\ -\psi_2 & 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -\psi_{q-1} & 0 & \cdots & \cdots & 0 & 1\\ -\psi_q+\phi_q & \phi_{q-1} & \cdots & \cdots & \phi_2 & \phi_1\end{bmatrix},$$

$$\boldsymbol{\alpha}=(\psi_1,\psi_2,\cdots,\psi_{q-1},\psi_q)^{\mathrm{T}},$$

$$\boldsymbol{\beta}=\left(0,0,\cdots,0,\sum_{j=q+1}^{p}\phi_jX_{k,q-j+1}\right),$$

而 $$\phi_j=\begin{cases}\varphi_j, & 0\leqslant j\leqslant p,\\ 0, & j>p.\end{cases}$$

当 $p\leqslant q$ 时，上式中 $\beta=0$. 初值取为

$$\hat{\boldsymbol{X}}_{k_0}^q=(\hat{X}_{k_0}(1),\hat{X}_{k_0}(2),\cdots,\hat{X}_{k_0}(q))^{\mathrm{T}},\qquad(31.14)$$

可由直接预报法计算得到它们. 在可逆性条件满足时，也可取 $\hat{X}_{k_0}^q=0$.

31.3 时间序列的新息实时预报

在许多实际问题中，特别是那些要求连续进行实时预报的问题，希望根据已获得的数据 $x_1, x_2, \cdots, x_t$，给出 x_{t+l}严格的线性最小方差预报，并希望有实时预报的递推算法. 本段介绍的新息实时预报，便是一个解决办法.

31.3.1 新息预报原理

新息预报原理　令

$$\mathscr{H}_k^1 = \left\{X; X = \sum_{i=1}^{k} a_i X_i, a_i \text{ 为实数}\right\},$$

根据 $X_k, X_{k-1}, \cdots, X_1$ 对 X_{k+l}所做的最小方差线性估计为

$$\hat{X}_{k+l,k} \equiv E(X_{k+l} | X_k, X_{k-1}, \cdots, X_1) \triangleq E(X_{k+l} | \mathscr{H}_k^1).$$

定义 31.3.1　最小方差线性估计中的一步预报误差

$$\varepsilon_k = X_k - \hat{X}_{k,k-1}, \quad k = 1, 2, \cdots.$$

所组成的序列$\{\varepsilon_k\}$为$\{X_k\}$的**新息序列**(new information series).

令 $\mathscr{E}_k^1 = \left\{\varepsilon; \varepsilon = \sum_{i=1}^{k} \delta_i \varepsilon_i, \delta_i \text{ 为实数}\right\}$，且设 $E\varepsilon_k X_j = 0, j < k$，$E\varepsilon_i\varepsilon_j = 0, i \neq j$，则 $\hat{X}_{k+l,k}$ 可表示为 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_k$ 的线性组合

$$\hat{X}_{k+l,k} = \sum_{j=1}^{k} J_{k+l,j} \varepsilon_j, \tag{31.15}$$

式中系数 $J_{k+l,j}$可由极小化 $E(X_{k+l} - \hat{X}_{k+l,k})^2$ 求出.

31.3.2 新息定理

定理 31.3.2　**新息定理**(new information theorem)　设$\{X_t\}$是满足 ARMA(p,q)模型的平稳零均值序列，记 $Q = \max(p,q)$，

令
$$Y_k=\begin{cases}X_k, & k\leqslant Q,\\ X_k-\sum_{i=1}^{p}\varphi_i X_{k-i}, & k>Q.\end{cases}\tag{31.16}$$

则新息序列$\{\varepsilon_k\}$满足下列递推关系：

$$\varepsilon_k=\begin{cases}Y_1=X_1, & k=1,\\ Y_k-\sum_{j=1}^{k-1}J_{kj}\varepsilon_j, & 1<k\leqslant Q,\\ Y_k-\sum_{j=k-q}^{k-1}J_{kj}\varepsilon_j, & k>Q.\end{cases}\tag{31.17}$$

其中
$$J_{kj}=\begin{cases}\left[r_{kj}(Y)-\sum_{i=1}^{j-1}J_{ji}J_{ki}r_{ii}(\varepsilon)\right](r_{jj}(\varepsilon))^{-1}, & \\ \qquad 1\leqslant k\leqslant Q,\ 1\leqslant j\leqslant k-1, & \\ \left[r_{kj}(Y)-\sum_{i=k-q}^{j-1}J_{ki}J_{ji}r_{ii}(\varepsilon)\right](r_{jj}(\varepsilon))^{-1}, & \\ \qquad k>Q,\ k-q\leqslant j\leqslant k-1. & \end{cases}\tag{31.18}$$

式中$r_{kj}(Y)=EY_kY_j$，$r_{jj}(\varepsilon)=E\varepsilon_j^2$，而

$$r_{kk}(\varepsilon)=\begin{cases}r_{11}(Y), & k=1,\\ r_{kk}(Y)-\sum_{j=1}^{k-1}J_{kj}^2r_{jj}(\varepsilon), & 1<k\leqslant Q,\\ r_{kk}(Y)-\sum_{j=k-q}^{k-1}J_{kj}^2r_{jj}(\varepsilon), & k>Q.\end{cases}\tag{31.19}$$

31.3.3 新息预报公式

公式 31.3.3 新息预报公式

$$\varepsilon_{k+l}=\begin{cases}X_{k+l}-\sum_{j=1}^{k+l-1}J_{k+l,j}\varepsilon_j, & k+l\leqslant Q,\\ Y_{k+l}-\sum_{j=k-q}^{k+l-1}\varphi_jX_{k+l-j}-\sum_{j=k+l-q}^{k+l-1}J_{k+l,j}\varepsilon_j, & k+l>Q.\end{cases}$$

上式经调整并在线性空间 $\mathscr{H}_k^1$ 上投影后，便得 X_k 的 l 步实时预报公式为

$$\hat{X}_{k+l,k}=\begin{cases}\sum_{j=1}^{k}J_{k+l,j}\varepsilon_j, & k+l\leqslant Q,\\ -\sum_{j=1}^{p}\varphi_j X_{k+l-j,k}+\sum_{j=k+l-q}^{k}J_{k+l,j}\varepsilon_j, & k+l>Q\end{cases}\tag{31.20}$$

当 $l-j\leqslant 0$ 时， $\hat{X}_{k+l-j,k}\equiv X_{k+l-j}$.

要注意，J_{kj} 只与 X_t 的模型参数值有关，而与它的样本值无关，ε_k 还与 X_t 的样本值 $X_1,X_2,\cdots,X_k$ 有关.

31.3.4 新息预报的程序设计

新息预报的计算过程如下：

(1) 假设在新息预报之前，根据已获得的有限长度的样本序列，建立了 ARMA(p,q)模型，并由(31.16)产生了长为 N 的序列 $\{Y_k\}$. 设 L 为预报的最大步长，随着 Y_k 的序贯输入，新息定理中的各种数字是依如下顺序进行计算：

$$\begin{array}{l}Y_1\\ \downarrow\\ \varepsilon_1\end{array}\to r_{11}(\varepsilon)\to J_{21}\to\begin{array}{l}Y_2\\ \downarrow\\ \varepsilon_2\end{array}\to r_{22}(\varepsilon)\to\begin{array}{l}J_{33}\\ J_{32}\end{array}\to\begin{array}{l}Y_3\\ \downarrow\\ \varepsilon_3\end{array}\to r_{33}(\varepsilon)\to\begin{array}{l}J_{41}\\ J_{42}\\ J_{43}\end{array}\to\cdots$$

其中计算初值为

$$\begin{cases}\varepsilon_1=Y_1=X_1,\\ r_{11}(\varepsilon)=r_0(X)=\dfrac{1}{N}\sum_{k=1}^{N}X_k^2.\end{cases}\tag{31.21}$$

(2) $r_{kj}(Y)$的计算

设 $k>j$，

$$r_{kj}(Y)=EY_ky_j=\begin{cases}EX_kX_j=r_{k-j}(X), & k\leqslant Q,\\ EY_kX_j=\varphi(B)r_{k-j}(X), & j\leqslant Q<k,\\ EY_ky_j=\varphi(B)\varphi^{-1}(B)r_{k-j}(X), & Q<j\leqslant k.\end{cases}\tag{31.22}$$

其中 $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - \varphi_2 B^{-2} - \cdots - \varphi_p B^{-p}$，算子 B 作用在 $r_{k-j}(X)$ 的下标 $(k-j)$ 上. 具体计算时，$r_{k-j}(X)$ 由它的样本估计值 $\frac{1}{N}\sum\limits_{l=1}^{N-(k-j)} x_l x_{l+(k-j)}$ 代替，当 $k \leqslant Q$ 时，r_{kj} 依下对角线计算较为方便. 而 $k>Q$ 时，按上对角线计算(如下表).

$$
\begin{array}{llllllll}
k=Q+1 & r_{Q+1,Q+1-q} & r_{Q+1,Q+2-q} & \cdots & & r_{Q+1,Q} & r_{Q+1,Q+1} & \\
k=Q+2 & & r_{Q+2,Q+2-q} & & \ddots & r_{Q+2,Q} & r_{Q+2,Q+1} & r_{Q+2,Q+2} \\
 & & & \ddots & \ddots & & \ddots & \ddots \\
k=Q+q+1 & & & & r_{Q+q+1,Q+1} & r_{Q+q+1,Q+2} & &
\end{array}
$$

$k>Q$ 时，$r_{kj}(Y)$的计算规律

(3) J_{kj}的计算

新息递推公式中，J_{kj}的计算工作量最大，计算过程最复杂. 分三种情形计算 J_{kj}.

1) $k \leqslant Q$

每计算一个 ε_k，需要用到 $J_{k1}, J_{k2}, \cdots, J_{k,k-1}$(31.17). 而计算 J_{kj}($j>1$)时，又要用到 $J_{j1}, J_{j2}, \cdots, J_{j,j-1}$ 和 $J_{k1}, J_{k2}, \cdots, J_{k,j-1}$(31.18). 计算量随 k 的增加而增加，而每计算一行 J_{kj}，$j=1,2,\cdots,k-1$，就要用到前面全部的 J_{ij}($j=1,2,\cdots,i-1$，$i=2,3,\cdots,k-1$)值，在 $2 \leqslant k \leqslant Q$ 范围内，$\{J_{kj}\}$元素的总个数为$\frac{1}{2}Q(Q-1)$.

2) $Q<k \leqslant N$

计算每一个 ε_k，只要计算 $J_{k,k-q}, \cdots, J_{k,k-1}$这 q 个值(31.17)，计算量不随 k 的增加而增加；计算每一个 J_{kj}，需依次用到 $J_{j,k-q}$，$\cdots$，$J_{j,j-1}$和 $J_{k,k-q}, \cdots, J_{k,j-1}$(31.18). 将计算 $J_{k,k-q}, \cdots, J_{k,k-1}$所依次用到的全体 J_{ki}与 J_{ji}排列起来，也是一个三角阵，阵中元素个数为$\frac{1}{2}q(q-1)$.

3) $N<k \leqslant N+L$

此时 $\varepsilon_k \equiv 0$. 为计算 $X_{k,N}$，只要知道 $J_{k,k-q}, \cdots, J_{k,N}$就够了. J_{kj}

的计算同(2)，但随 k 的增加，需要计算的 J_{kj} 个数是递减的，如图所示的倒置直角梯形.

$$\begin{matrix} J_{N+1,N+1-q} & J_{N+1,N+2-q} & \cdots & \cdots & J_{N+1,N} \\ & J_{N-2,N+2-q} & \cdots & \cdots & J_{N+2,N} \\ & & \cdots & \cdots & \cdots \\ & & J_{N+L,N+L-q} & \cdots & J_{N+L,N} \end{matrix}$$

31.4 非平稳时间序列的预报

在许多实际问题中，由量测直接得到的随机序列 $\{X_t, t\in T\}$ 并不平稳，而是呈现出明显的趋势性(trend property)和季节性(seasonal property). 这时 $\{X_t, t\in T\}$ 是非平稳过程. 取

$$X_t = \mu_t + Y_t, \tag{31.23}$$

其中 μ_t 是 X_t 的均值函数，它可以用多项式、指数函数、对数函数等描述；Y_t 为零均值平稳时间序列，可用 ARMA(p,q)模型拟合. 处理这类问题的方法是通过某些处理，剔除 μ_t 的变化趋势；或者是具体求出 μ_t 的拟合形式.

31.4.1 ARIMA(p,d,q)模型的预报

设 $\{Z_t, t\in T\}$ 为满足 ARIMA(p,d,q)非平稳时间序列：

$$\varphi(B)(1-B)^d Z_t = \theta(B)\varepsilon_t,$$

l 步预报为

$$\hat{Z}_k(l) = \sum_{i=0}^{d-1} C_i^{l+i-1} \nabla^i Z_k + \sum_{j=1}^{l} C_{d-1}^{l-j+d-1} \hat{X}_k(j), l > 0.$$

式中 $\nabla^i Z_t = X_t$, $(t>d)$, X_t 是平稳 ARMA(p,q)序列.

若 $d=1$,

$$\hat{Z}_k(l) = Z_k + \sum_{j=1}^{l} \hat{Z}_k(j), \quad l > 0. \tag{31.24}$$

若 $d=2$,

$$\hat{Z}_k(l) = Z_k + l(Z_k - Z_{k-1}) + \sum_{j=1}^{l}(l+1-j)\hat{X}_k(j), \quad l > 0. \tag{31.25}$$

31.4.2 ARIMA$(p,d,q)\times(P,D,Q)$模型的预报

设$\{Z_t, t\in T\}$满足乘积型季节模型(23.10.2),$D>0$. 记

$$X_t = \nabla^d \nabla_s^D Z_t,$$

则 X_t 为 ARMA(p,q)序列,$\hat{X}_k(l)$满足预报差分方程(31.2.5),而(31.9)是此方程的解. 记

$$\tilde{\varphi}(B) = \nabla^d \nabla_s^D \varphi(B).$$

由于 $\hat{X}_k(l)$具有线性性质,所以

$$\hat{X}_k(l) = \nabla^d \nabla_s^D \hat{Z}_k(l), \quad k > d + D, \quad l \geqslant 0. \tag{31.26}$$

其中 k 固定,算子∇,∇_s都作用在 l 上,于是当 $k>d+D$ 时,

$$\tilde{\varphi}(B)\hat{Z}_k(l) = 0, \quad l > q. \tag{31.27}$$

这个差分方程的通解形式为

$$\hat{Z}_k(l) = \sum_{i=1}^{p+d+Ds} b_i^{(k)} f_i(l), \quad l > q + (p + d + D \times s) \tag{31.28}$$

其中 $f_i(l)$$(1\leqslant i\leqslant p+d+D\times s)$是(31.28)的 $p+d+Ds$ 个线性无关解. 它们可以表示为 l 的指数函数、幂函数或三角函数的乘积形式. $b_i^{(k)}$ 可由初值 $\hat{Z}_k(q)$,$\hat{Z}_k(q-1)$,…,$\hat{Z}_k(q-p-d-Ds+1)$确定,将其代入(31.28),所得即为 Z_t 的预报函数.

若 $D=0$,则(31.28)为

$$\hat{Z}_k(l) = h_0^{(k)} + h_1^{(k)} l + \cdots + h_{d-1}^{(k)} l^{d-1} + \sum_{i=1}^{p} b_i^{(k)} f_i(l), \quad l > q - (p + d). \tag{31.29}$$

若 $p=0,D=0$,则 $\hat{Z}_k(l)$ 只剩(31.29)式中的多项式部分.

31.4.3 组合模型

设$\{X_t,t\in T\}$满足(31.23)式

$$X_t = \mu_t + Y_t,$$

μ_t 为序列的确定性部分,Y_t 为零均值平稳随机过程,用 ARMA(p,q)拟合. 用这两部分叠合起来共同描述 X_t,称为组合模型. μ_t 的表达形式可以为:

(1) 线性趋势(linear trend)

先用统计方法拟合回归直线,再对残差建立 ARMA(p,q)模型,最后把分别估计得到的两部分参数作为初值,用非线性最小二乘估计出组合模型参数.然后进行预报.

(2) 指数趋势(exponential trend)

$$X_t = \sum_{j=1}^{r} A_j e^{k_j t} + Y_t, \tag{31.30}$$

其中 k_j,A_j 为实数,$\{Y_t\}$为 ARMA 序列,预报为

$$\hat{X}_t(l) = \sum_{j=1}^{r} A_j e^{k_j t} + \hat{Y}_t(l), \quad l > 1. \tag{31.31}$$

(3) 周期趋势(periodic trend)

如果序列本身不仅含有指数增长或衰减趋势,而且还显示出一定规律的周期性起伏,则可用如下模型

$$X_t = \sum_{j=1}^{L} R_j e^{r_j t} + \sum_{j=1}^{K} B_j e^{b_j t}\sin(j\omega t + \psi_j) + Y_t, \tag{31.32}$$

其中$\{Y_t\}$是 ARMA(p,q)序列,L 为序列中含有的指数趋势项数,K 相应于周期趋势项数,R_j,r_j 为实数$(j=\overline{1,L})$. ω 为基频,单位为弧度/秒,B_j 和 ψ_j 分别表示周期性趋向的振幅和相位,$e^{b_j t}$控制周期项的增长或衰减. 由于估计 ψ_j 困难,因此将(31.32)改写为

$$X_t = \sum_{j=1}^{L} R_j e^{r_j t} + \sum_{j=1}^{K} B_j e^{b_j t}[c_j\sin(j\omega t) + \sqrt{1-c_j^2}\cos(j\omega t)]$$

$$+Y_t,\tag{31.33}$$

其中 $c_j=\cos\phi_j$. 用非线性最小二乘法估计出未知参数后，把确定性部分的预报值与 ARMA(p,q)序列的预报值相叠加，便得到观察序列$\{X_t,t\in T\}$的预报值.

31.5 时间序列的频域预报

设$\{X_t,t\in T\}$是离散时间**正则过程**(regular process)，$f(\lambda)=\frac{1}{2\pi}|\varphi(e^{-i\lambda})|^2$ 是它的谱密度，$X_t=\int_{-\pi}^{\pi}e^{it\lambda}d\zeta(\lambda)$ 是它的谱表现，而 $X_t=\sum_{s=-\infty}^{t}c(t-s)\zeta_0(s)$ 是它的滑动和过程形式，其中$\sum_{t=0}^{\infty}|c(t)|^2<\infty$.

定理 31.5.1 设平稳列 $X_t,t\in T$ 有前述谱密度，则根据观察 $X_s,s\leqslant t$，对于 $X_{t+\tau},\tau>0$ 的值提前 τ 时的最优线性预报为

$$\hat{X}_t(\tau)=\int_{-\pi}^{\pi}e^{i(t+\tau)\lambda}\frac{\varphi_\tau(e^{-i\lambda})}{\varphi(e^{-i\lambda})}d\zeta(\lambda)$$

其中 $\varphi(e^{-i\lambda})=\sum_{t=0}^{\infty}c(t)e^{-it\lambda},\quad \varphi_\tau(e^{-i\lambda})=\sum_{t=\tau}^{\infty}c(t)e^{-it\lambda}.$

预报误差

$$\begin{aligned}\sigma_\tau^2&=E|X_{t+\tau}-\hat{X}_t(\tau)|^2=\sum_{s=0}^{\tau-1}|c(s)|^2\\&=2\pi\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln f(\lambda)d\lambda\right\}\sum_{s=0}^{\tau-1}|d_s|^2,\end{aligned}$$

其中 d_s 决定于关系式

$$\exp\left\{\frac{1}{\pi}\sum_{n=1}^{\infty}z^n\int_{-\pi}^{\pi}e^{in\lambda}\ln f(\lambda)d\lambda\right\}=\sum_{n=0}^{\infty}d_nz^n.\tag{31.34}$$

特别

$$\sigma_1^2=2\pi\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln f(\lambda)d\lambda\right\},\tag{31.35}$$

即$\frac{\sigma_1^2}{2\pi}$是谱密度(连续)的几何均值.

例 31.5.2 设$\{X_t,t\in T\}$是广义马尔可夫过程(即 AR(1)过程),它的相关函数为$B(t)=\sigma^2 e^{-\alpha|t|}$,$\alpha>0$,其中$T=\{0,\pm1,\pm2,\cdots\}$,谱密度

$$f(\lambda)=\frac{\sigma^2}{2\pi}\frac{1-\beta^2}{|1-\beta e^{-i\lambda}|^2},\quad \beta=e^{-\alpha}.$$

提前时间τ的最优线性预报为

$$\hat{X}_t(\tau)=\int_{-\pi}^{\pi}e^{i\lambda t}e^{-\alpha\tau}d\zeta(\lambda)=e^{\alpha\tau}X_t,$$

其中$\zeta(\lambda)$是对应于过程$\{X_t,t\in T\}$的谱过程.预报误差为

$$\sigma_\tau^2=\sigma^2(1-e^{-2\alpha\tau}).$$

定理 31.5.1 可扩展到一般有理谱密度的情形,即

$$f(\lambda)=\frac{1}{2\pi}\Big|\sum_{k=0}^{q}\theta_k e^{-ik\lambda}\Big|^2\Big/\Big|\sum_{l=0}^{p}\varphi_l e^{-il\lambda}\Big|^2,\quad -\pi\leqslant\lambda\leqslant\pi \tag{31.36}$$

的情形.下一定理对另一种形式的谱密度,给出预报结论.为叙述简洁,设$t=0$.

定理 31.5.3 设$X_s,s\leqslant0$是平稳列,其谱密度为

$$f(\lambda)=f^*(e^{i\lambda}),$$

$$f^*(z)=\left|\frac{B_0}{A_0}\right|^2\left(\prod_{k=1}^{q}(z-b_k)(1-\bar{b}_k z)\right.$$

$$\left.\Big/\prod_{l=1}^{p}(z-a_l)(1-\bar{a}_l z)\right)z^{p-q},$$

其中$0<|a_l|<1$,$0<|b_k|<1$,

当$p\geqslant q$时,确定一个$p-1$次多项式$Q(z)$,使

$$z^{\tau+p-q}\prod_{k=1}^{q}(z-b_k)-Q(z).$$

以$a_l,l=1,2,\cdots,p$为根;

当 $p<q$ 时，确定一个 $q-1$ 次多项式 $Q(z)$，使

$$z^{\tau}\prod_{k=1}^{q}(z-b_k)-Q(z).$$

以 $a_l, l=1,\cdots,p$ 为根并有 $q-p$ 个零根，令

$$\Phi_{\tau}(z)=\begin{cases}Q(z)\Big/\left[z^{p-q}\prod\limits_{k=1}^{q}(z-b_k)\right], & p\geqslant q,\\ Q(z)\Big/\prod\limits_{k=1}^{q}(z-b_k), & p<q.\end{cases}$$

则最优预报

$$\hat{X}_0(\tau)=\int_{-\pi}^{\pi}\Phi_{\tau}(\mathrm{e}^{\mathrm{i}\lambda})\mathrm{d}\zeta(\lambda).$$

若记预测误差为 σ_{τ}^2，则

$$\lim_{\tau\to\infty}\sigma_{\tau}^2=\int_{-\pi}^{\pi}f(\lambda)\mathrm{d}\lambda=R(0).$$

例 31.5.4 设平稳列 $X(s),s\in T$ 有谱密度 $f(\lambda)=c|\mathrm{e}^{\mathrm{i}\lambda}-b|^2$，$0<|b|<1$. 求最优预报 $\hat{X}_0(\tau)$.

解 由谱密度式知 $p=0,q=1(>p)$，故为MA(1)模型，且复函数 $f^*(z)=c(z-b)(1-bz)z^{-1}$. 因此应找 $q-1=0$ 阶多项式 $Q(z)$，即常数 α，使 $z^{\tau}(z-b)-\alpha$ 以 $z=0$ 为根. 令 $\tau=0$，由 $z^0(z-b)|_{z=0}-\alpha=0$ 得 $\alpha=-b$. 故

$$\Phi_{\tau}(z)=-b/(z-b)=\frac{-b}{z}\sum_{k=0}^{\infty}(b/z)^k,\ |z|>1.$$

$$\begin{aligned}\hat{X}_0(\tau)&=\int_{-\pi}^{\pi}\Phi_{\tau}(\mathrm{e}^{\mathrm{i}\lambda})\mathrm{d}\zeta(\lambda)\\&=-\sum_{k=1}^{\infty}\int_{-\pi}^{\pi}\mathrm{e}^{-\mathrm{i}k\lambda}b^k\mathrm{d}\zeta(\lambda)\\&=-\sum_{k=1}^{\infty}b^kX(-k).\end{aligned}$$

$\tau>0$ 时，$z^{\tau}(z-b)-\alpha=0$ 要有零根，只有 $\alpha=0$. 从而 $\Phi_{\tau}(z)\equiv 0$，故 $\hat{X}_0(\tau)\equiv 0$. 这一事实，与 MA(1)一步截尾性相合：间隔两步以上的随机变量间是无关的.

本段介绍的频域预报，都是在谱密度已知条件下作出的，这些方法归功于雅格龙(Яглом).

31.6 滤波

31.6.1 引言

设时刻 $t\in T_0\subset T$，观察过程为

$$X(t) = S(t) + \theta(t),$$

它是有用信号 $S(t)$ 与噪声 $\theta(t)$ 之和，其中 $\{S(t),t\in T\}$ 是平稳且不相关过程. 要求把噪声 $\theta(t)$ 滤出，提取 $S(t)$，即找出一个依赖于过程 $\{X(t),t\in T_0\}$ 的泛函

$$Y(t^*) = g_{t^*}(X(t),t \in T_0),t^* \in T_0,\text{使}$$

$$E\|S(t^*) - Y(t^*)\|^2 \leqslant E\|S(t^*) - Y_1(t^*)\|^2,t^* \in T_0, \tag{31.37}$$

其中 $Y_1(t^*)$ 是依赖于 $\{X(t),t\in T_0\}$ 的任意泛函.

称(31.37)左侧期望值为滤波误差(filter error).

在20世纪40年代，科尔莫戈罗夫、克莱因和维纳等人做了大量工作并取得了许多重要成果. 其理论和应用在控制和通讯等技术领域中受到重视. 但是，作为信号估计的维纳滤波是基于对线性系统受平稳噪声干扰的频域分析，计算比较复杂，不便于实时处理，因而应用范围受到了很大限制. 直到60年代初，由卡尔曼提出了最优线性递推滤波(optimum linear recursive filtering)或卡尔曼滤波，它由一套递推关系式构成，便于用计算机在线实时计算. 同时把应用范围扩展到非平稳噪声干扰的线性时变系统，基本上克服

了维纳滤波的上述几个主要缺点，因此得到了广泛的应用.

31.6.2 线性系统的卡尔曼滤波

设线性离散随机系统为

$$\begin{cases} X(t+1)=A(t)X(t)+\xi(t+1), \\ Y(t)=C(t)X(t)+\eta(t). \end{cases} \quad t\in T_d \qquad (31.38)$$

其中 $X(t)\in R^n$，是系统的状态，$Y(t)\in R^m$ 是系统的输出，$\boldsymbol{A}(t)\boldsymbol{C}(t)$是适当维数的矩阵，$\xi(t)$，$\boldsymbol{\eta}(t)$，$t\in T_d$ 分别是 n 维和 m 维白噪声过程. 设 t 时刻有量测数据集合

$$\boldsymbol{Y}^t \triangleq (Y(t_0),Y(t_0+1),\cdots,Y(t)), \quad t\in T_d \qquad (31.39)$$

要讨论的问题是利用 $\boldsymbol{Y}^t$ 去估计 $X(t)$.

公式31.6.1 基本卡尔曼滤波方程 设线性系统如(31.38)所示，并假设$\{\xi(t)\}$、$\{\eta(t)\}$相互独立，且

$$\begin{cases} \xi(t)\sim N(0,R_1(t)),\ t=t_0+1,t_0+2,\cdots,R_1(t)\geqslant 0, \\ \eta(t)\sim N(0,R_2(t)),t=t_0,\ t_0+1,\cdots,R_2(t)>0. \end{cases} \qquad (31.40)$$

初态 $X(t_0)$与噪声独立，且

$$X(t_0)\sim N(\overline{X}(t_0),\boldsymbol{P}(t_0)) \qquad (31.41)$$

那么对估计误差的任意容许损失函数，基于 $\boldsymbol{z}(t)=\boldsymbol{Y}^t=[Y^T(t_0),Y^T(t_0+1),\cdots,Y^T(t)]^{\mathrm{T}}$，对状态 $X(t)$的最小均方误差估计满足如下递推方程：

$$\begin{cases} \hat{X}(t+1|t+1)=\boldsymbol{A}(t)\hat{X}(t|t)+\boldsymbol{K}(t+1)[Y(t+1) \\ \qquad\qquad -\boldsymbol{C}(t+1)\boldsymbol{A}(t)\hat{X}(t|t)], \\ \hat{X}(t_0|t_0)=E\{X(t_0)\}+\boldsymbol{K}(t_0)[Y(t_0)-\boldsymbol{C}(t_0)E(X(t_0))], \end{cases} \qquad (31.42)$$

其中 $\hat{X}(t|t)\triangleq E\{X(t)|z(t)\}$，$z(t_0)=Y(t_0)$，$\boldsymbol{K}(t+1)$ 是卡尔曼滤波增益阵，满足

$$\begin{cases} \boldsymbol{K}(t+1)=[\boldsymbol{A}(t)\boldsymbol{\Sigma}(t|t)\boldsymbol{A}^{\mathrm{T}}(t)+\boldsymbol{R}_1(t+1)] \\ \qquad \cdot\boldsymbol{C}^{\mathrm{T}}(t+1)\{\boldsymbol{C}(t+1)[\boldsymbol{A}(t)\boldsymbol{\Sigma}(t|t)\boldsymbol{A}^{\mathrm{T}}(t) \\ \qquad +\boldsymbol{R}_1(t+1)]\boldsymbol{C}^{\mathrm{T}}(t+1)+\boldsymbol{R}_2(t+1)\}^{-1} \\ \boldsymbol{K}(t_0)=\boldsymbol{P}(t_0)\boldsymbol{C}^{\mathrm{T}}(t_0)[\boldsymbol{C}(t_0)\boldsymbol{P}(t_0)\boldsymbol{C}^{\mathrm{T}}(t_0)+\boldsymbol{R}_2(t_0)]^{-1} \end{cases} \qquad (31.43)$$

这里 $\boldsymbol{\Sigma}(t|t) \triangleq E\{[X(t)-\hat{X}(t|t)][X(t)-\hat{X}(t|t)]^{\mathrm{T}}\}$，即估计误差的协方差阵，满足

$$\begin{cases}\boldsymbol{\Sigma}(t+1|t+1)=[\boldsymbol{I}_n-\boldsymbol{K}(t+1)\boldsymbol{C}(t+1)] \\ \qquad\qquad \cdot [\boldsymbol{A}(t)\boldsymbol{\Sigma}(t|t)\boldsymbol{A}^{\mathrm{T}}(t)+\boldsymbol{R}_1(t+1)] \\ \boldsymbol{\Sigma}(t_0|t_0)=[\boldsymbol{I}_n-\boldsymbol{K}(t_0)\boldsymbol{C}(t_0)]\boldsymbol{P}(t_0)\end{cases} \tag{31.44}$$

其中 $\boldsymbol{I}_n$ 表示 n 阶单位矩阵. 若 $\hat{X}(t+1|t)$ 为基本变量，则递推公式为

$$\begin{cases}\hat{X}(t+1|t)=\boldsymbol{A}(t)\hat{X}(t|t-1)+\boldsymbol{K}^*(t)[Y(t)-\boldsymbol{C}(t) \\ \qquad\qquad \hat{X}(t|t-1)], \\ \hat{X}(t_0|t_0-1)=E\{X(t_0)\}=\overline{X}(t_0).\end{cases} \tag{31.45}$$

$$\boldsymbol{K}^*(t)=\boldsymbol{A}(t)\boldsymbol{\Sigma}(t|t-1)\boldsymbol{C}^{\mathrm{T}}(t)[\boldsymbol{C}(t)\boldsymbol{\Sigma}(t|t-1)\boldsymbol{C}^{\mathrm{T}}(t) \\ +\boldsymbol{R}_2(t)]^{-1}. \tag{31.46}$$

其中 $\boldsymbol{\Sigma}(t|t-1)$ 是预报误差的协方差阵，满足

$$\begin{cases}\boldsymbol{\Sigma}(t+1|t)=[\boldsymbol{A}(t)-\boldsymbol{K}^*(t)\boldsymbol{C}(t)]\boldsymbol{\Sigma}(t|t-1)\boldsymbol{A}^{\mathrm{T}}(t) \\ \qquad\qquad +\boldsymbol{R}_1(t+1), \\ \boldsymbol{\Sigma}(t_0|t_0-1)=\operatorname{cov}(X(t_0),X(t_0))=\boldsymbol{P}(t_0).\end{cases} \tag{31.47}$$

31.6.3 有平稳噪声干扰时线性定常系统稳态滤波

考虑线性定常系统

公式31.6.2

$$\begin{cases}X(t+1)=\boldsymbol{A}X(t)+\xi(t) \\ Y(t)=\boldsymbol{C}X(t)+\eta(t)\end{cases}$$

假设 $\xi(t)\sim N(0,R_1)$，$\eta(t)\sim N(0,R_2)$ 为相互独立的高斯白噪声过程，$\boldsymbol{A}$，$\boldsymbol{C}$，$\boldsymbol{R}_1$，$\boldsymbol{R}_2$均为常数阵，其它假设与基本卡尔曼滤波相同.

由(31.46)知，(31.47)的一步提前预报误差的协方差阵满足

$$\boldsymbol{\Sigma}(t+1|t)=\boldsymbol{A}[\boldsymbol{\Sigma}(t|t-1)-\boldsymbol{\Sigma}(t|t-1) \\ \boldsymbol{C}^{\mathrm{T}}[\boldsymbol{C}\boldsymbol{\Sigma}(t|t-1)\boldsymbol{C}^{\mathrm{T}}+\boldsymbol{R}_2]^{-1}\cdot\boldsymbol{C}\boldsymbol{\Sigma}(t|t-1)]\boldsymbol{A}^{\mathrm{T}} \\ +\boldsymbol{R}_1, t\geqslant t_0. \tag{31.48}$$

定义31.6.3 如果存在一个n阶方阵$\overline{\Sigma}$是(31.48)的解，则称此卡尔曼滤波器是**定常的**(deterministic)；如果 $\Sigma(t|t-1)\xrightarrow[n\to\infty]{}\overline{\Sigma}$，则称此卡尔曼滤波器是**渐近定常的**(asymptotically deterministic).

如此滤波器是定常的，则预报的卡尔曼增益阵也是常阵，

$$K^{*}=A\overline{\Sigma}C^{\mathrm{T}}[C\overline{\Sigma}C^{\mathrm{T}}+R_2]^{-1}. \qquad (31.49)$$

定理31.6.4 对于有平稳噪声干扰的线性定常系统，如果该系统是渐近稳定的，对于初态的任意协方差阵，卡尔曼滤波器的一步提前预报的误差协方差阵 $\Sigma(t|t-1)$ 收敛于一个常阵$\overline{\Sigma}$，$\overline{\Sigma}$满足代数里卡蒂(Riccati)方程：

$$\overline{\Sigma}=A[\overline{\Sigma}-\overline{\Sigma}C^{\mathrm{T}}[C\overline{\Sigma}C^{\mathrm{T}}+R_2]^{-1}C\overline{\Sigma}]A^{\mathrm{T}}+R_1, \qquad (31.50)$$

而且滤波器本身是渐近稳定的，如果该线性系统不是渐近稳定的，但只要$[A,C]$是能观测的，且 $R_1=GG^{\mathrm{T}}$ 定义的 G 与 A 构成一个完全能稳定对，则滤波器本身是渐近稳定的，而且 $\Sigma(t|t-1)$ 收敛于$\overline{\Sigma}$.

例31.6.5 α-β-γ 滤波器 设对连续时间信号 $S(t)$，作周期为Δ的采样，当Δ很小时，有近似

$$S((k+1)\Delta)-S(k\Delta)=\Delta S'(k\Delta)+\frac{\Delta^2}{2}S''(k\Delta).$$

令 $x(k)=(S(k\Delta),S'(k\Delta),S''(k\Delta))^{\mathrm{T}}$，则加速度受随机干扰时，有状态方程

$$x(k+1)=Ax(k)+\sigma\xi(k),$$

其中 $\sigma=(0,0,1)^{\mathrm{T}}$，$\xi(k)$为高斯白噪声，服从 $N(0,\gamma_1^2)$，$A=(a_{ij})_{3\times3}$，$a_{ii}=1$，$a_{12}=a_{21}=\Delta$，$a_{13}=\dfrac{\Delta^2}{2}$，其余全为0. 量测

$$y(k)=Cx(k)+\eta(k),$$

$\eta(k)$也为高斯白噪声，服从 $N(0,\gamma_2^2)$，且与$\{\xi(k)\}$独立，于是有定常系统(31.6.2)，令

$$\Sigma_{ij}(k|k-1)=\mathrm{cov}\{x_i(k)-\hat{x}_i(k|k-1),x_j(k)-\hat{x}_j(k|k-1)\},$$

则由(31.42)、(31.43)，增益与滤波公式为

$$\boldsymbol{K}(k)=\begin{bmatrix}\Sigma_{11}(k|k-1)\\ \Sigma_{21}(k|k-1)\\ \Sigma_{31}(k|k-1)\end{bmatrix}(\Sigma_{11}(k|k-1)+\gamma_2^2)^{-1},$$

$$\hat{x}(k|k)=(\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{C})\boldsymbol{A}\hat{x}(k-1|k-1)+\boldsymbol{K}(k)y(k).$$

初始时刻 $k_0\to-\infty$时，波波器进入稳态，于是令 $\boldsymbol{K}=(\alpha,\beta,\gamma)^{\mathrm{T}}$，则

$$\hat{x}(k|k)=\boldsymbol{\Phi}\hat{x}(k-1|k-1)+\boldsymbol{K}y(k),$$

其中

$$\boldsymbol{\Phi}=\begin{bmatrix}1-\alpha & (1-\alpha)\Delta & (1-\alpha)\dfrac{\Delta^2}{2}\\ -\beta & 1-\beta\Delta & (1-\beta\Delta/2)\Delta\\ -\gamma & -\gamma\Delta & 1-\gamma\Delta^2/2\end{bmatrix}.$$

此即 α-β-γ 滤波器. 设由(31.50)解得 $\overline{\boldsymbol{\Sigma}}$, 则

$$\boldsymbol{K}=(\overline{\Sigma}_{11},\overline{\Sigma}_{21},\overline{\Sigma}_{31})^{\mathrm{T}}/(\overline{\Sigma}_{11}+\gamma_2).$$

32 排队论

32.1 引言

32.1.1 排队问题的例子

在公共服务事业的许多部门如商店、医院、银行和公共汽车站都存在排队现象. 在商店中顾客陆续到来，营业员逐个接待每个顾客，在顾客人数较多时就要排队等待，这样顾客和营业员构成了一个**排队系统**(queueing system). 在系统中的顾客数描述了系统的状态. 随着时间的推移，一些顾客得到服务后离去了，又有一些新的顾客到来，所以排队系统是随时间变化的动态系统. 顾客到来的时间有随机性，对顾客的服务时间也有随机性，因此研究排队问题自然会使用随机过程的工具，所以排队论也称为**排队过程**(queueing process). 在排队模型中服务是一个重要因素，因此排队论也称为**随机服务系统**(stochatic service system).

在排队论中的术语**顾客**(customer)、**服务员**(server)其含义可以是各式各样的. 例如在医院门诊部前就诊的患者就是顾客，应诊的医生就是服务员. 在生产车间的一道工序上，待加工的工件就是顾客，加工设备或操作工人就是服务员. 再如，港口的装卸设施作为服务员，那么到港装货或卸货的轮船就是顾客. 排队论有广泛的应用.

32.1.2 排队模型的三要素

每个排队系统都是由以下三个基本要素结合而成.

(1) **输入过程**(input process)或**到达过程**(arrival process).指顾客到来时间的规律,通常用到达间隔时间来描述.设第 n 个到来的顾客是在时刻 t_n 到来的,令 $s_n=t_{n+1}-t_n$(约定 $t_0=0$),$\{s_n\}$是到达间隔时间序列.通常假定$\{s_n\}$是独立同分布的随机变量.

$X(t)$表示在区间$(0,t)$内到来的顾客数,随机过程$\{X(t),t\geqslant 0\}$叫做排队系统的**输入过程**.可以证明,间隔时间独立同分布且为指数分布等价于 $X(t)$是泊松过程,即 $X(t)$有独立平稳增量,并且

$$P(X(t+s)-X(s)=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},t>0,\lambda>0$$

另外,顾客可以是单个到达也可以是成批到达,而每批的顾客数可以是固定的也可以是随机的.

(2) **服务机构**(service mechanism).服务员可以是一个或多个.对各个顾客的服务时间通常假定是独立同分布的随机变量.对顾客也可以成批进行服务.以上所说是指一个服务站的情形.另外也可以由多个服务站构成网络,顾客按一定规则接受几个站的服务.

(3) **排队规则**(queue discipline).有多种型式的排队规则.当一个顾客到来时服务员均不空闲,此顾客就排队等待一直到得到服务后离去,这称为**等待制**(waiting system).反之当一个顾客到来时服务员均不空闲,此顾客立即离去不再到来,这称为**损失制**(loss system).介于等待制与损失制之间的是**限制队长的排队**(system limiting queue size),即先规定一个正数 m,顾客到来时如果系统中的顾客数不足 m 个他便加入系统,否则便离去不再来.服务员在服务完一个顾客后如果系统中有顾客在等待,服务员便按一定的规则从中选定一个顾客开始对他服务,选定的规则有:**先到先服务**(first-come-first-served),**后到先服务**(last-come-first-served),**随机地指定**(service in random order)和**按某种优先规则**(service according to priority).

32.1.3 排队系统主要数量指标

排队论主要研究排队系统的三个主要数量指标：等待时间，忙期和队长.

(1) **等待时间**(waiting time)　指顾客自到来至他开始接受服务这段时间. 在实际问题中顾客总是希望等待时间尽可能少，等待时间的长短是服务质量的一个方面.

(2) **忙期**(busy period)　忙期指服务员由空闲转为开始工作到下一次出现空闲这段时间. 忙期反映了服务员工作的劳累程度.

(3) **队长**(queue length)　指系统中的顾客数，包括正在服务的和排队等待的. 队长是排队系统的一个很直观的指标，它和等待时间、忙期都有联系.

排队论研究这几个主要指标，力图把它们的概率分布解析地表示出来. 任意时刻的队长分布、第 n 个顾客的等待时间分布，第 n 个忙期分布，这些称为排队系统的**瞬时解**. 当时间参数趋向无穷时这些指标的极限分布，称为排队系统的**稳态解**. 求瞬时解比求稳态解要困难得多，瞬时解的表达式也复杂得多. 但从实际应用的角度看，稳态解似乎是更重要的.

32.1.4 排队论模型的记号

这种记号由三个符号组成，中间用两个斜杠"/"分隔开. 例如 $M/G/1$ 代表一种排队模型. 其中第一个符号表示顾客到来间隔时间的分布类型，第二个符号表示服务时间分布类型. 具体规定是：M 代表指数分布，D 代表定常分布，G 代表一般分布，E_k 代表 k 阶埃尔朗(Erlang)分布，PH 代表**位相型**(phase type)分布. 第三个符号是一个正整数，代表服务员个数. 这样 $M/G/1$ 排队的意思是，其模型中顾客到来间隔时间服从指数分布，对顾客的服务时间服从任意给定的分布，服务站只有一个服务员.

32.2 生灭过程排队模型

32.2.1 生灭过程的一个极限定理

设$\{N(t),t\geqslant 0\}$是齐次生灭过程，则其$\boldsymbol{Q}$矩阵具有以下形式

$$\boldsymbol{Q}=\begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -\lambda_2-\mu_2 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -\lambda_3-\mu_3 & \lambda_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{bmatrix},$$

$$\lambda_i>0,\mu_i>0. \tag{32.1}$$

此外，我们假定矩阵$\boldsymbol{Q}$是有界的，即存在正数C对一切i均有$\lambda_i+\mu_i\leqslant C$.

考虑关于$(x_0,x_1,x_2,\cdots)$的下列方程组.

$$\begin{cases}\lambda_0x_0=\mu_1x_1, \\ (\lambda_i+\mu_i)x_i=\lambda_{i-1}x_{i-1}+\mu_{i+1}x_{i+1}, \end{cases} \quad i\geqslant 1. \tag{32.2}$$

取$\tilde{x}_0=1$，得到方程组的一组解：

$$\tilde{x}_i=\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}, \quad i\geqslant 1; \quad \tilde{x}_0=1.$$

构成级数

$$\sum_{i=0}^{\infty}\tilde{x}_i=1+\sum_{i=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}, \tag{32.3}$$

该级数可能收敛也可能发散.

定理 32.2.1 设生灭过程的密度矩阵$\boldsymbol{Q}$是有界的，则极限$\lim\limits_{t\to\infty}P(N(t)=j)=\pi_j,j\geqslant 0$存在并且与初始分布无关. 此外还有

(1) 若级数(32.3)发散，则一切$\pi_j=0$，方程组(32.2)不存在概率分布解，生灭过程$N(t)$不存在平稳分布.

(2) 若级数(32.3)收敛，则一切$\pi_j>0$，$(\pi_0,\pi_1,\pi_2,\cdots)$是方程

组(32.2)的惟一概率分布解,而且是过程 $N(t)$ 的惟一平稳分布.此外还有

$$\begin{cases} \pi_0 = \left[1 + \sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\right]^{-1}, \\ \pi_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\pi_0, \quad j \geqslant 1 \end{cases} \tag{32.4}$$

此时生灭过程 $N(t)$ 是正常返且遍历的,又 $\pi_j, j\geqslant 0$ 为过程 $N(t)$ 的**稳态概率**.

32.2.2 $M/M/1$ 排队系统

$M/M/1$ 排队是单个服务员的排队系统,顾客到来间隔时间服从参数为 λ 的指数分布,服务时间服从参数为 μ 的指数分布,且有隐含假定:顾客间隔到来时间是独立同分布的,服务时间也是独立同分布的并且独立于输入过程,排队规则是等待制的.今后把这些隐含假定作为是不言而喻的,不再特别申明.

在时刻 t 系统中的顾客数记为 $N(t)$,由于指数分布的无后效性,$N(t)$ 是齐次可列马尔可夫链,转移概率 $P_{ij}(t)\equiv P(N(t+s)=j|N(s)=i), i, j\geqslant 0$. 对充分小的 $h>0, P_{i\,i+1}(h)=P$(在 $(0,h)$ 时间内到来一个顾客并且正在服务的尚未服务完,或者在 $(0,h)$ 内到来 $k\geqslant 2$ 个顾客并且已服务完 $k-1$ 个顾客)$=(1-e^{-\lambda h})+o(h)=\lambda h+o(h), i\geqslant 0$ 类似的有

$$P_{i\,i-1}(h) = \mu h + o(h), \quad i \geqslant 1,$$
$$P_{ij}(h) = o(h), \qquad |i-j| > 1.$$

由以上各式可立即得出 $N(t)$ 的密度矩阵 $\boldsymbol{Q}$:将(32.1)中 λ_i, μ_i 分别易为 λ, μ.

利用(32.1)容易得出,当 $\rho\equiv\frac{\lambda}{\mu}\geqslant 1$ 时 $N(t)$ 不是遍历的.此时对于任意正整数 k 均有

$$\lim_{t\to\infty} P(N(t) \leqslant k) = \sum_{j=0}^{k} \pi_j = 0.$$

粗略地说，这表明当时间充分长时系统中顾客充分多的概率接近于 1. 称这样的排队系统是不稳定的. 而当 $\rho<1$ 时 $N(t)$遍历，队长过程 $N(t)$的稳态概率由(32.4)立即得出为

$$\pi_j = (1-\rho)\rho^j, j \geqslant 0. \tag{32.5}$$

在稳定状态下的平均队长是

$$L \equiv \sum_{k=1}^{\infty} k\pi_k = \frac{\rho}{1-\rho}. \tag{32.6}$$

在时刻 t 系统中排队等待的顾客数记为 $N_q(t)$，

$$N_q(t) = \max(N(t)-1, 0),$$

$N_q(t)$按分布收敛意义下的极限记为 N_q，则

$$P(N_q = k) = \begin{cases} 1-\rho^2, & k = 0, \\ (1-\rho)\rho^{k+1}, & k \geqslant 1. \end{cases} \tag{32.7}$$

N_q 的数学期望是

$$L_q \equiv \sum_{k=1}^{\infty} kP(N_q = k) = \frac{\rho^2}{1-\rho}, \tag{32.8}$$

顾客在系统的停留时间是他的等待时间与服务时间之和. 在节 32.5 我们将说明第 n 个顾客的等待时间与停留时间分别依分布收敛于随机变量 $\widetilde{W}_q$ 与 $\widetilde{W}$(分别称为在稳定状态下顾客的等待时间与停留时间).

公式 32.2.2 $\widetilde{W}_q$ 与 $\widetilde{W}$ 的分布

$$P(\widetilde{W}_q \leqslant t) = \begin{cases} 1-\rho, & t = 0, \\ 1-\rho e^{-\mu(1-\rho)t}, & t > 0, \end{cases}$$

$$P(\widetilde{W} \leqslant 1) = 1 - e^{-\mu(1-\rho)t}, \quad t \geqslant 0,$$

$$E(\widetilde{W}_q) = \frac{\rho}{\mu(1-\rho)}, E(\widetilde{W}) = \frac{1}{\mu(1-\rho)},$$

及 $$L = \lambda W, \quad L_q = \lambda W_q.$$

32.2.3 其它生灭过程排队系统

1. $M/M/n$ 排队系统

n 个服务员的排队系统，顾客到来间隔时间服从参数为 λ 的指数分布，服务时间服从参数为 μ 的指数分布. 仍用 $N(t)$ 表示队长过程，这也是一个生灭过程，$P_{ij}(t)$ 表示它的转移概率，对充分小的 $h>0$，

$$P_{i\,i+1}(h)=\lambda h+o(h),\quad i\geqslant 0,$$

$$P_{i\,i-1}(h)=\min(i\mu,n\mu)+o(h),\quad i\geqslant 1.$$

于是作为密度矩阵(32.1)式中的 λ_i、μ_i 有

$$\lambda_i=\lambda,\quad i\geqslant 0;\quad \mu_i=\min(i\mu,n\mu),i\geqslant 1.$$

根据(32.1)，易见当且仅当 $\rho=\dfrac{\lambda}{m\mu}<1$ 时过程 $N(t)$ 是遍历的，按(32.4)得队长过程的稳态概率是

$$\pi_0=\left[1+\sum_{k=0}^{n-1}\frac{(n\rho)^k}{k!}+\frac{(n\rho)^n}{n!(1-\rho)}\right]^{-1},$$

$$\pi_k=\frac{(n\rho)^k}{k!}\pi_0,\quad 1\leqslant k<n,$$

$$\pi_k=\frac{n^n\rho^k}{n!},\quad k\geqslant n.$$

在稳定状态下，系统的平均顾客数是

$$L=\sum_{k=1}^{\infty}k\pi_k=\left[\sum_{i=1}^{n-1}\frac{(n\rho)^i}{(i-1)!}+\frac{(n\rho)^n}{n!}\frac{\rho+n(1-\rho)}{(1-\rho)^2}\right]\pi_0.$$

2. 有限源排队系统

在织布车间一个工人看管 n 台自动织布机. 当机器发生故障时就停下来，经工人修复和适当处置后再开动起来继续工作. 设每台机器能正常工作的持续时间是独立同分布的，服从参数为 λ 的指数分布，工人每修复一台机器所需时间也是独立同分布的，服从参数为 μ 的指数分布，它们还独立于机器正常工作持续时间. 把

正在修复与等待修复的机器作为顾客，把工人作为服务员，便构成一个排队系统. 顾客得到服务离去后会再来，作为顾客源的只是这 n 台机器，这类排队系统叫**有限源排队**(finite-source queue). 用 $N(t)$ 表示时刻 t 系统中的顾客数，这是一个有限状态生灭过程. 对充分小的 $h>0$

$$P_{i\ i+1}(h)=(n-i)\lambda h+o(h),\quad 0\leqslant i\leqslant n-1,$$

$$P_{i\ i-1}(h)=\mu h+o(h),\quad 1\leqslant i\leqslant n.$$

因此，密度矩阵可以 $(n-i)\lambda,\mu$ 分别易(32.1)中 λ_i,μ_i 得到.

有限状态生灭过程总是遍历的. 稳态概率 $\pi_i=\lim\limits_{t\to\infty}P(N(t)=i)$，仍按(32.4)计算，但只须将级数的求和上限改为 n，这样得出

$$\pi_0=\left[\sum_{i=0}^{n}\left(\frac{\lambda}{\mu}\right)^i\frac{n!}{(n-i)!}\right]^{-1},$$

$$\pi_k=\left(\frac{\lambda}{\mu}\right)^k\frac{n!}{(n-k)!}\pi_0,\quad 1\leqslant k\leqslant n.$$

32.3 马尔可夫链排队系统

32.3.1 马尔可夫链的一个极限定理

设 $\{N(t),t\geqslant 0\}$ 是齐次可列马尔可夫链，有标准转移矩阵，假定它是不可约的，其密度矩阵为 $\boldsymbol{Q}=[q_{ij}]$，其中，$-q_{ii}=\sum\limits_{j\neq i}q_{ij}>0$. 还假定 $\boldsymbol{Q}$ 是有界的，即存在正数 C，对一切 i 均满足 $-q_{ii}\leqslant C$.

定理 32.3.1 对于马尔可夫链 $N(t)$，在上述假定条件下，极限 $\lim\limits_{t\to\infty}P(N(t)=j)=\pi_j$，$j\geqslant 0$，存在并且与初始分布无关. 此外还有下列两种情况之一成立：

(1) 一切 $\pi_j=0$，过程 $N(t)$ 不存在平稳分布.

(2) 一切 $\pi_j>0$，$\pi\equiv(\pi_0,\pi_1,\pi_2,\cdots)$ 是 $N(t)$ 唯一的平稳分布，并且它是稳态平衡方程

$$\sum_{i=0}^{\infty} x_i q_{ij} = 0, \quad j \geqslant 0 \tag{32.9}$$

唯一的概率分布解. 此时 $N(t)$是遍历的.

32.3.2 成批服务排队系统

考虑一个服务员的排队系统,顾客到来间隔时间及服务时间分别服从参数为 λ, μ 的指数分布. 对顾客成批的进行服务. 这样的**成批服务**(bulk-service)排队系统其简写记号是 $M/M^{[X]}/1$. 通常一次服务 r 个顾客,在服务员出现空闲而系统中有顾客但少于 r 个时,有两种规则:

(1) 服务员等待,直到来齐 r 个顾客时便开始服务.

(2) 把系统中现有的顾客作为一批进行服务,在服务过程中新到来的顾客排队等待.

$N(t)$表示时刻 t 的队长,这是一个马尔可夫链. 可以证明对上述两种规则当且仅当 $\rho \equiv \frac{\lambda}{r\mu} < 1$ 时 $N(t)$是遍历的.

先讨论第(1)种规则. 对任意小的 $h>0$

$$\begin{aligned} &P_{i\,i+1}(h) = \lambda h + o(h), && i \geqslant 0, \\ &P_{i\,i-r}(h) = \mu h + o(h), && i \geqslant r, \\ &P_{ij}(h) = o(h), && \text{其它 } i \neq j. \end{aligned} \tag{32.10}$$

由此可得出密度矩阵 $\boldsymbol{Q}$.

试分析过程 $N(t)$的一个样本. 当系统在某一时刻处于状态 i 时, 下一次发生转移只能转移到 i 或 $i-r$(当 $i \geqslant r$ 时). 参照(32.10)式,称 i 到 $i+1$ 的转移率是 λ,i 到 $i-r$ 的转移率是 μ. 把系统的这种转移关系表示为一个有向图,以状态作为结点,直接转移关系表示为一个有向弧,在弧上标出转移率,这种图叫**状态转移率图**(state-transition-rate diagram).

成批服务系统的状态转移率图(图 32.1)表示 i 到 j 的转移率

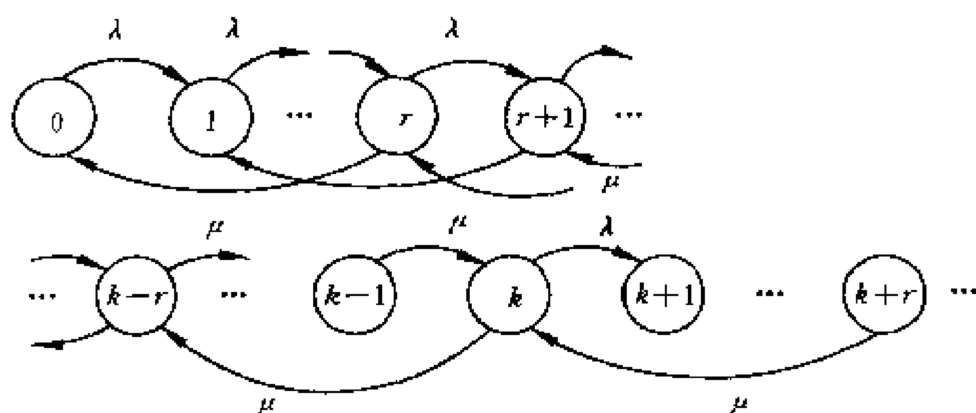


图 32.1

是矩阵 Q 中的元素 q_{ij}，$i \neq j$. 元素 q_{ii} 按条件 $q_{ii} = -\sum\limits_{j \neq i} q_{ij}$ 确定. 这样参照转移图可以写出矩阵 Q. 更进一步的参照转移图还可直接写出稳态概率 ν 满足的平衡方程(32.9)，其规则是：第 k 个方程为

$$\sum_j (k \text{ 到 } j \text{ 的转移率})\pi_k = \sum_i (i \text{ 到 } k \text{ 的转移率})\pi_i.$$

对于我们的排队系统，按此规则便得出

$$\begin{cases} \lambda\pi_0 = \mu\pi_r, \\ \lambda\pi_j = \lambda\pi_{j-1} + \mu\pi_{r+j}, \quad 1 \leqslant j \leqslant r-1, \\ (\lambda+\mu)\pi_j = \lambda\pi_{j-1} + \mu\pi_{r+j}, \quad j \geqslant r. \end{cases} \tag{32.11}$$

定义母函数 $Q(z) \equiv \sum\limits_{j=0}^{\infty} \pi_j z^j$，$|z| < 1$，将(32.11)式两端乘以 z^j，对 j 求和整理后得

$$Q(z) = \frac{\mu(1-z^r)\sum\limits_{j=0}^{r-1}\pi_j z^j}{\lambda z^{r+1} - (\lambda+\mu)z^r + \mu}.$$

在 $\rho<1$ 时可以证明上式分母的多项式恰好有一个大于 1 的实根 z_0，并且 $Q(z)$ 可以化为

$$Q(z) = \frac{1}{r}\left(1 - \frac{1}{z_0}\right)\left(1 - \frac{z}{z_0}\right)^{-1}\left(\sum_{j=0}^{r-1} z^j\right),$$

将$\left(1-\frac{z}{z_0}\right)^{-1}$展开为幂级数后相乘，再对照$Q(z)$的定义得

$$\begin{cases}\pi_j=\frac{1}{r}\left[1-\left(\frac{1}{z_0}\right)^{j+1}\right], & 0\leqslant j<r,\\ \pi_j=\frac{\lambda}{r\mu}(z_0-1)\left(\frac{1}{z_0}\right)^{j+1-r}, & j\geqslant r.\end{cases}\tag{32.12}$$

在稳定状态下的平均队长是

$$\sum_{k=1}^{\infty}k\pi_k=Q'(1)=\frac{1}{z_0-1}+\frac{r-1}{2}.\tag{32.13}$$

下面讨论成批服务的第(2)种排队规则. 不难得出系统的转移率图如图(32.2). 参照转移率图得出稳态概率π_j满足的方程组是

$$\lambda\pi_0=\mu(\pi_1+\pi_2+\cdots+\pi_r),$$

$$(\lambda+\mu)\pi_j=\lambda\pi_{j-1}+\mu\pi_{j+r},\quad j\geqslant 1.$$

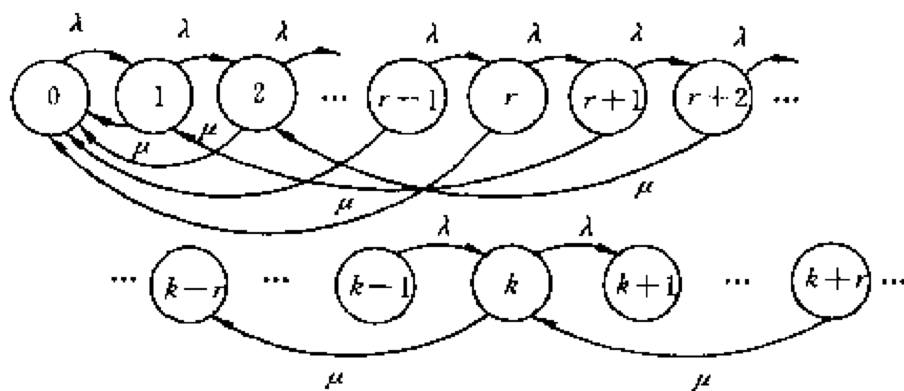


图 32.2

用类似的方法得到母函数$Q(z)$的表达式为

$$Q(z)=\frac{\mu\sum_{j=0}^{r}\pi_jz^j-(\lambda+\mu)\pi_0z^r}{\lambda z^{r+1}-(\lambda+\mu)z^r+\mu}$$

$$=\frac{(z_0-1)}{(z_0-2)}=\left(1-\frac{1}{z_0}\right)\sum_{j=0}^{\infty}\left(\frac{2}{z_0}\right)^j.$$

由此得出队长的稳态概率

$$\pi_j = \left(1 - \frac{1}{z_0}\right)\left(\frac{1}{z_0}\right)^j, \quad j \geqslant 0$$

在稳态下的平均队长是

$$\sum_{j=0}^{\infty} j\pi_j = \frac{1}{z_0 - 1}.$$

32.3.3 $E_r/M/1$ 排队系统

$E_r/M/1$ 排队系统有一个服务员，服务时间服从参数为 μ 的指数分布，顾客到达间隔时间服从参数为 λ 的 r 阶埃尔朗分布，分布密度是

$$\frac{\lambda(\lambda t)^{r-1}}{(r-1)!}e^{-\lambda t}, \quad t \geqslant 0.$$

熟知，r 个独立参数为 λ 的指数分布随机变量之和服从上述埃尔朗分布. 因此一个埃尔朗变量的发展进程是相连接的 r 个指数变量进程之和. 每实现 r 个指数变量的一个进程我们视为到来一个顾客. 而服务看作是成批进行的，每批 r 个，排队规则采用第 1 种形式. 这个成批排队系统在时刻 t 的队长记为 $N(t)$，它的初始分布满足条件：若 j 不是 r 的整倍数则 $P(N(0)=j)=0$. 前已说明当 $\rho = \frac{\lambda}{r\mu} < 1$，$N(t)$ 遍历其稳态概率 π_j 由(32.12)给出.

对于 $E_r/M/1$ 排队系统，在时刻 t 的队长为 $X(t)$，于是有 $X(t) = [N(t)/r]$，方括号表示整数部分. 记 $X(t)$ 的稳态概率 $P_k = \lim_{t\to\infty} P(X(t)=k)$，从而有 $P_k = (\pi_{k_r} + \pi_{k_{r+1}} + \cdots + \pi_{k_{r+r-1}})$，利用(32.12)式得：

$$P_0 = 1 - \rho, \tag{32.14}$$

$$P_k = \rho(z_0^r - 1)z_0^{-kr}, \quad k \geqslant 1,$$

其中 z_0 的意义同前. $E_r/M/1$ 排队系统在稳态下的平均队长是

$$\sum_{k=1}^{\infty} kP_k = \frac{1}{r(z_0 - 1)}. \tag{32.15}$$

32.4 $M/G/1$ 排队系统

32.4.1 嵌入链

$M/G/1$ 排队系统有一个服务员，对各顾客的服务时间是独立同分布的随机变量，其分布函数 $B(t)$ 已知，服务时间的均值为 $\frac{1}{\mu}$，方差为 σ^2. 顾客到来的间隔时间是独立同分布的，服从参数为 λ 的指数分布，服务时间与输入过程独立.

假定在起始时刻 $t=0$，刚好服务完一个顾客. 令 q_n 表示第 n 个被服务的顾客服务完毕离开系统的瞬时所留下的队长. v_n 表示在为第 n 个顾客服务时间内新到来的顾客数，于是

$$q_{n+1}=\begin{cases}q_n-1+v_{n+1}, & \text{若 } q_n>0,\\ v_{n+1}, & \text{若 } q_n=0.\end{cases}$$

容易看出 $\{v_n\}$ 是一列独立同分布的随机变量. 第 n 次服务时间记为 u_n，由于指数分布的无后效性，v_n 只与 u_n 的大小有关而与起点无关

$$P(v_n=k\mid u_n=t)=\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t},$$

$$a_k\equiv P(v_n=k)=\int_{0-}^{\infty}\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}\mathrm{d}B(t).$$

又可看出 v_{n+1} 与 $(q_0,q_1,\cdots,q_n)$ 独立. 所以 $\{q_n\}$ 是一个齐次马尔可夫链. 称它为这个排队系统的**嵌入链**，而顾客离开系统的时刻叫**嵌入点**. 转移概率 $P_{ij}=P(q_{n+1}=j\mid q_n=i)$，对此有

$$P_{ij}=P(v_{n+1}=j+1-i)=\begin{cases}a_{j+1-i}, & j+1\geqslant i\geqslant 1,\\ 0, & j+1<i \text{ 且 } i\geqslant 1.\end{cases}$$

$$P_{0j}=P(v_{n+1}=j)=a_j,\quad j\geqslant 0.$$

转移概率矩阵 $\boldsymbol{P}=[P_{ij}]$ 具有以下形式

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_3 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}.$$

可见 $\boldsymbol{P}$ 是不可约周期的. 在马尔可夫链理论中有以下极限定理.

定理 32.4.1 对于不可约非周期可列马尔可夫链, $P_{ij}^{(n)}$ 是它的 n 步转移概率, 则极限

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j, \quad j \geqslant 0$$

存在与 i 无关并且与初始分布无关. 此外还有以下两种情况之一成立

(1) 一切 $\pi_j=0$, 马尔可夫链不存在平稳分布.

(2) 一切 $\pi_j>0$, $(\pi_0,\pi_1,\cdots)$ 是马尔可夫链的唯一平稳分布, 并且是稳态平衡方程组

$$\sum_{i=0}^{\infty} x_i P_{ij} = x_j, \quad j \geqslant 0$$

的唯一概率分布解. 此时马尔可夫链是遍历的.

对于 $M/G/1$ 排队系统的嵌入链 $\{q_n\}$, 可以证明, 当且仅当 $\rho=\dfrac{\lambda}{\mu}<1$ 时是遍历的. 以下假设 $\rho<1$, 我们来研究在稳态下的队长和等待时间.

32.4.2 队长

令 $P_j^{(n)} = P(q_n = j)$, 从而 $P_j^{(n)} = \sum_{i=0}^{\infty} P(q_0 = i) P_{ij}^{(n)}$ 由定理 32.4.1 得

$$\lim_{n\to\infty} P_j^{(n)} = \pi_j, \quad j \geqslant 0.$$

则 $Q^{(k)}(0)$ 表示 $Q(z) \triangleq \sum_{k=0}^{\infty} \pi_k z^k$ 在 $z = 0$ 的 k 阶导数.

$$\pi_k = \frac{1}{k!}Q^{(k)}(0), \quad k \geqslant 0. \tag{32.16}$$

$(\pi_0, \pi_1, \cdots)$应满足平衡方程组，即

$$\pi_0 a_j + \sum_{k=1}^{j+1} \pi_k a_{j-1-k} = \pi_j, \quad j \geqslant 0.$$

将上式两端乘以 z^j，对 j 相加整理得

$$Q(z) = \frac{\pi_0(1-z)f(z)}{f(z)-z}, \quad |z| < 1.$$

其中 $f(z)$ 为 $\{a_k\}$ 的母函数，令 $z \to 1_-$，对上式取极限可求得 $\pi_0 = 1-\rho$，代入上式成为

$$Q(z) = \frac{(1-\rho)(1-z)f(z)}{f(z)-z}, \quad |z| < 1. \tag{32.17}$$

这样我们求出了母函数 $Q(z)$ 的表达式. 原则上说，利用(32.16)，(32.17)就可求出稳态概率 π_k.

在稳态下的平均队长 $L = \sum_{k=1}^{\infty} k\pi_k = Q'(1)$. 将(32.16)两端求导数，令 $z \to 1_-$ 取极限得

$$L = \lim_{z \to 1_-} Q'(z) = \rho + \frac{\lambda^2\sigma^2 + \rho^2}{z(1-\rho)}. \tag{32.18}$$

32.4.3 等待时间

分布函数 $B(x)$ 的拉普拉斯-斯蒂尔切斯(Laplace-Stieltjes)变换记为 $B^*(s)$，其定义是

$$B^*(s) = \int_{0_-}^{\infty} e^{-sx} dB(x), \quad s > 0.$$

假定排队规则是先到先服务的. 可以证明，在 $\rho<1$ 的条件下，第 n 个顾客的等待时间依分布收敛于随机变量 w——这是稳定状态下顾客的等待时间，其分布函数记为 $W(x)$.

在稳态下的队长 q_n 记为 q. 顾客离开时刻的队长是在这个顾

客的等待时间与服务时间内到来的顾客数，而同一个顾客的等待时间与服务时间是独立的，所以有

$$P(q=k)=\int_{0_-}^{\infty}\int_{0}^{\infty}\frac{[\lambda(x+y)]^k}{k!}e^{-\lambda(x+y)}\mathrm{d}W(x)\mathrm{d}B(y),$$

$$\begin{aligned}Q(z)&=\sum_{k=0}^{\infty}P(q=k)z^k\\&=\int_{0}^{\infty}\int_{0_-}^{\infty}e^{-\lambda(1-z)(x+y)}\mathrm{d}W(x)\mathrm{d}B(y)\\&=W^*(\lambda(1-z))B^*(\lambda(1-z)).\end{aligned}$$

将此式同(32.17)联系，注意到式中的 $f(z)$ 就是 $B^*(\lambda(1-z))$，并令 $s=\lambda(1-z)$，得

$$B^*=\frac{(1-\rho)^s}{\lambda B^*(s)-\lambda+s},\quad s>0,\qquad (32.19)$$

此式称为扑拉切克-辛钦(Pollaczek-Хинчин)公式. 令 $s\to\infty$ 得顾客不要等待的概率

$$P(w=0)=W(0^+)=1-\rho.$$

考虑稳态下顾客平均等待时间 $E(w)$

$$\begin{aligned}E(q)&=\sum_{k=1}^{\infty}kP(q=k)=\int_{0_-}^{\infty}\int_{0_-}^{\infty}\lambda(x+y)\mathrm{d}W(x)\mathrm{d}B(y)\\&=\lambda E(w)+\rho.\end{aligned}$$

联系(32.18)式得

$$E(w)=\frac{\lambda^2\sigma^2+\rho^2}{2\lambda(1-\rho)},\qquad (32.20)$$

平均服务时间记为 $E(v)=\frac{1}{\mu}$，于是

$$\frac{E(w)}{E(v)}=\frac{\rho}{2(1-\rho)}(1+\mu^2\sigma^2).\qquad (32.21)$$

这个指标是有意义的. 顾客为了得到一定的服务时间需另外再花费一定的等待时间，而这个时间之比反映了服务效率. $\frac{E(w)}{E(v)}$ 愈小，表示服务效率愈高.

32.5 $G/M/1$ 排队系统

32.5.1 队长

$G/M/1$ 排队系统有一个服务员，服务时间是独立同分布的服从参数为 λ 的指数分布. 顾客到来的间隔时间也是独立同分布的，分布函数为 $A(x)$，均值为 $\frac{1}{\mu}$. 服务时间与输入过程独立.

第 n 个顾客在时刻 τ_n 到来，在这个瞬间他看到系统中的顾客数记为 q_n，ν_n 表示在时间区间 (τ_n, τ_{n+1}) 中服务员实际完成的服务次数，于是有

$$q_{n+1} = q_n + 1 - \nu_n$$

在已知 q_n 的条件下，ν_n 与 $(q_0, q_1, \cdots, q_{n-1})$ 独立，因而 $\{q_n\}$ 是马氏链.

注 32.5.1 记号

$$b_k \equiv \int_{0_-}^{\infty} \frac{(\mu x)^k}{k!} \mathrm{e}^{-\mu x} \mathrm{d}A(x), \quad k \geqslant 0,$$

$$r_k = 1 - \sum_{j=0}^{k} b_j, \quad k \geqslant 0.$$

b_k 的概率意义是：在一次到来时间间隔中，服务员不间断地连续工作完成 k 次服务的概率.

$$P(\nu_n = j | q_n = i) = \begin{cases} b_j, & 若\ i + 1 > j \geqslant 0, \\ r_{j-1}, & 若\ i + 1 = j \geqslant 0, \\ 0, & 若\ j < 0\ 或\ j > i + 1. \end{cases}$$

记一步转移概率 $P_{ij} = P(q_{n+1} = j | q_n = i)$，则得

$$P_{ij} = \begin{cases} b_{i+1-j}, & 若\ i + 1 \geqslant j > 0, \\ r_i, & 若\ j = 0, \\ 0, & 若\ j > i + 1. \end{cases} \tag{32.22}$$

可以证明，当且仅当 $\rho=\frac{\lambda}{\mu}<1$ 时 $\{q_n\}$ 是遍历的. 以下均假定 $\rho<1$.

嵌入链 q_n 的稳态概率 $\pi_j=\lim P(q_n=j)$ 满足的平衡方程是

$$\sum_{j=0}^{\infty} r_j\pi_j = \pi_0,$$

$$\sum_{i=0}^{\infty} b_i\pi_{j-1+i} = \pi_j, j \geqslant 1.$$

现来求 π_j. 首先令

$$f(x) = \sum_{k=0}^{\infty} b_k x^k, g(x) = f(x) - x, \quad 0 \leqslant x \leqslant 1,$$

在 $(0,1)$ 内 $g''(x)=f''(x)>0$，$g(x)$ 是凸函数. 又 $g(0)=b_0>0$，$g(1)=0$，$g'(1)=\frac{1}{\rho}-1<0$，所以 $g(x)$ 在 $(0,1)$ 内有唯一的零点，记为 z_0.

可以验证稳态平衡方程的解是

$$\pi_k = (1 - z_0)z_0^k, \quad k \geqslant 0. \tag{32.23}$$

在嵌入链意义下稳态时，系统的平均队长是

$$\sum_{k=1}^{\infty} k\pi_k = \frac{z_0}{1 - z_0}. \tag{32.24}$$

32.5.2 等待时间

$w_q^{(n)}$ 表示第 n 个到来顾客的等待时间. 在 $q_n=0$ 时 $w_q^{(n)}=0$，在已知 $q_n=i>0$ 的条件下 $w_q^{(n)}$ 是连续服务 i 个顾客花费的时间，因而服从 i 阶埃尔朗分布. 所以有

$$P(w_q^{(n)} \leqslant t | q_n = i) = \int_0^t \frac{\mu(\mu x)^{i-1}}{(i-1)!} e^{-\mu x} dx,$$

$$P(w_q^{(n)} \leqslant t) = \sum_{i=1}^{\infty} P(w_q^{(n)} \leqslant t | q_n = i) P(q_n = i) + P(q_n = 0),$$

$$\lim_{n\to\infty} P(w_q^{(n)} \leqslant t) = (1 - z_0)z_0 \int_0^t e^{-\mu(1-z_0)x} dx + (1 - z_0)$$

$$= 1 - z_0 e^{-\mu(1-z_0)t}, t \geqslant 0.$$

在稳定状态下顾客的等待时间记为 $\tilde{w}_q$，上式表明它的分布函数是

$$P(\tilde{w}_q \leqslant t) = \begin{cases} 1 - z_0, & t = 0 \\ 1 - z_0 e^{-\mu(1-z_0)t}, & t > 0. \end{cases} \qquad (32.25)$$

$$E(\tilde{w}_q) = \frac{z_0}{\mu(1-z_0)}$$

在稳态情况下，顾客在系统的停留时间为 $\tilde{w} = \tilde{w}_q + v$，其中 $\tilde{w}_q, v$ 是同一个顾客的等待时间与服务时间因而它们是独立的，可求得

$$P(\tilde{w} \leqslant t) = 1 - e^{-\mu(1-z_0)t}, (t \geqslant 0) \qquad (32.26)$$

$$E(\tilde{w}) = \frac{1}{\mu(1-z_0)}$$

特别地如果到来间隔时间服从指数分布，那么前面定义的函数 $f(x)$ 即

$$f(x) = \sum_{k=0}^{\infty} b_k x^k = \frac{\lambda}{\lambda + \mu - \mu x}, \quad 0 \leqslant x \leqslant 1$$

方程 $f(x) - x = 0$ 在 $(0,1)$ 内的唯一解是 $z_0 = \frac{\lambda}{\mu} = \rho$，这说明了我们对 $M/M/1$ 排队曾提到的结果.

33 可靠性

33.1 引言

可靠性是系统最重要的指标之一,它定量地反映系统在一定的工作环境下,一定的时间内完成其预定功能的能力.通常用寿命来表示系统能正常工作的时间,用失效或故障表示系统处于不能正常工作的状态.**可靠性理论**(reliability theory)研究系统的失效机理及其寿命的概率规律.通过这种研究和分析把所获信息反馈到系统的设计、制造、运行、维修和管理等部门,以便使这些部门提出措施保证系统最大限度地完成其功能.

可靠性数学是可靠性理论的基础之一.它通过建立数学模型,对寿命数据做统计分析,以及在各种资源约束下对系统可靠性进行优化等方面的研究,为提高可靠性提供定量依据.

33.2 可靠性数量指标

定义 33.2.1 记系统的寿命为 X,是一个非负随机变量,相应的分布函数为 $F(t)$,称

$$\overline{F}(t) = 1 - F(t) = P(X > t), \quad t \geqslant 0. \tag{33.1}$$

为**可靠度函数**(reliability function). $\overline{F}(t)$ 为系统在时刻 t 以前完成其预定功能的概率.系统的平均寿命为

$$\mu = EX = \int_0^\infty t\mathrm{d}F(t) = \int_0^\infty \overline{F}(t)\mathrm{d}t. \tag{33.2}$$

若 X 有相应的密度函数 $f(t)$,则**失效率函数**(failure rate

function)定义为

$$r(t) = f(t)/\bar{F}(t), \quad t \geqslant 0, \quad \left\{\frac{0}{0} = 0\right\}. \qquad (33.3)$$

失效率函数的概率意义为

$$r(t)\Delta t \approx P(X \leqslant t + \Delta t \mid x > t).$$

即,$r(t)$表示系统到 t 正常时的条件失效密度. 失效率与可靠度之间有如下关系

$$\bar{F}(t) = \bar{F}(0)\exp\left\{-\int_0^t r(u)\mathrm{d}u\right\}, \quad t \geqslant 0. \qquad (33.4)$$

33.3 常用寿命分布

本段用 X 表示寿命,$F(t)$及 $f(t)$分别表示分布及密度函数,$r(t)$表示失效率函数.

33.3.1 指数分布

$$F(t) = 1 - \mathrm{e}^{-\lambda t}, \quad f(t) = \lambda \mathrm{e}^{-\lambda t}, \quad t \geqslant 0, \qquad (33.5)$$

$\lambda > 0$ 为参数, $EX = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}.$

$$r(t) = \lambda = \text{const}, \quad t \geqslant 0.$$

性质 33.3.1 指数分布的重要性质:无记忆性,$\forall\, x, y \geqslant 0$ $P(X > x + y \mid x > y) = P(X > x)$,或 $\bar{F}(x+y) = \bar{F}(x)\bar{F}(y)$.

定理 33.3.2 正随机变量 X 有指数分布,当且仅当 X 具有无记忆性(33.3.1).

33.3.2 韦布尔分布

$$F(t) = 1 - \mathrm{e}^{-(\lambda t)^\beta}, \quad f(t) = \lambda\beta(\lambda t)^{\beta-1}\mathrm{e}^{-\lambda t^\beta}, \quad t \geqslant 0, \qquad (33.6)$$

$\lambda, \beta > 0$,参数 λ, β 分别称为尺度与形状参数,$\beta = 1$ 时,即化为指数分布.

$$EX = \frac{1}{\lambda}\Gamma\left(1+\frac{1}{\beta}\right), V(X) = \frac{1}{\lambda^2}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right\},$$

这里 $\Gamma(\cdot)$ 为 Γ 函数.

$r(t)=\lambda\beta t^{\beta-1}, t\geqslant 0$，当 $\beta>1$ 时，$r(t)$ 递增；$\beta<1$ 时，递减；$\beta=1$ 时为常数.

33.3.3 Γ分布 $\Gamma(\alpha,\lambda;t)$

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \quad t > 0, \tag{33.7}$$

$\lambda,\alpha>0$ 参数，$EX=\frac{\alpha}{\lambda}$，$V(X)=\frac{\alpha}{\lambda^2}$.

当 $\alpha>1$ 时，$r(t)$ 递增；$\alpha<1$ 时，递减；$\alpha=1$ 时，$r(t)$ 为常数，化为指数分布.

33.3.4 对数正态分布

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t}\exp\left\{-\frac{1}{2\sigma^2}(\log t - \mu)^2\right\}, \quad t > 0. \tag{33.8}$$

$\sigma>0$，$-\infty<\mu<\infty$ 为参数.

$$EX = e^{\mu+\frac{1}{2}\sigma^2},$$

$$V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

失效率 $r(t)$ 先递增然后递减.

33.3.5 截尾正态分布

$$f(t) = \frac{1}{a\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\},$$

$$t > 0, \sigma > 0, -\infty < \mu < \infty, \tag{33.9}$$

其中 $a=\Phi(A)$，$A=\frac{\mu}{\sigma}$，$\Phi(t)$ 为 $N(0,1)$ 的分布函数.

$$EX = \mu + \frac{\sigma}{a}\varphi(A), V(X) = \sigma^2\left\{1 - \frac{A}{a}\varphi(A) - \frac{1}{a^2}\varphi^2(A)\right\},$$

其中 $\varphi(t)$为 $N(0,1)$的密度函数.

失效率函数 $r(t)$递增.

33.3.6 离散型寿命分布

二点分布、几何分布、二项分布、负二项分布、泊松分布等非负离散随机变量的分布都可用作寿命分布.

33.4 不可修复系统分析

本段中若无特别声明,则表明系统由 n 个独立元件组成,元件 i 的寿命为 x_i,分布为 $F_i(t)$,$i=1,\cdots,n$. 系统的寿命为 X,相应的分布为 $F(t)$.

33.4.1 串联和并联系统

若其中一个元件失效就引起系统失效,则称该系统为**串联系统**. 此时,

$$X=\min_{1\leqslant i\leqslant n}X_i,\quad \overline{F}(t)=\prod_{i=1}^{n}\overline{F}_i(t),\quad t\geqslant 0.$$

若所有元件失效,系统才失效,则称该系统为**并联系统**. 此时,

$$X=\max_{1\leqslant i\leqslant n}X_i,\quad F(t)=\prod_{i=1}^{n}F_i(t),\quad t\geqslant 0.$$

33.4.2 冷备系统

当 $t=0$ 时某元件开始运行,其余元件冷贮备(即不管元件贮备多久,工作时其寿命与新的一样). 一旦该元件失效,则由开关瞬时转换到贮备件. 系统失效发生在开关转换失效时,或备件用完时.

设开关使用一次成功的概率为 p,且各次使用互相独立. 记 N

为开关首次失效时刻，则

$$P(N=k)=p^{k-1}q,\ k=1,2,\cdots,p,\ q\geqslant 0,\ p+q=1.$$

$$X=\sum_{i=1}^{N}X_i\quad(\text{约定 } X_k=0,\text{当 } k>n \text{ 时}).$$

$$F(t)=\sum_{j=1}^{n}c_jF_1*\cdots*F_j(t),\quad t\geqslant 0.$$

其中 * 为卷积，$A*B(t)=\int_{[0,t]}B(t-u)\mathrm{d}A(u)$，

$$c_j=\begin{cases}p^{j-1}q, & j=1,\cdots,n-1,\\ p^{n-1}, & j=n.\end{cases}$$

特别当 X_i 同参数 λ 的指数分布时，

$$\overline{F}(t)=\mathrm{e}^{-\lambda t}\sum_{j=0}^{n-1}\frac{(p\lambda t)^j}{j!},t\geqslant 0.$$

$$EX=\frac{1}{\lambda q}(1-p^n).$$

33.4.3 热备系统

所谓热贮备是指在贮备过程中元件可能失效，或劣化.

为简单起见只讨论 $n=2$ 的热备系统. $t=0$ 时元件 1 开始运行，元件 2 热贮备. 若元件 1 失效时元件 2 未失效，则由开关转换使元件 2 立即转入工作. 否则，若元件 2 已在贮备中失效，则系统就失效. 假定开关转换完全可靠，记元件 i 的工作寿命为 X_i，$i=1,2$，元件 2 的贮备寿命为 Y. 则系统寿命为 $X=X_1+X_2I_{(X_1>Y)}$，这里 I_A 为 A 的示性函数.

33.5 单调关联系统

设系统由 n 个元件组成，记 x_i 为元件 i 的状态. $x_i=1$ 或 0，若元件 i 正常或失效. 记 $\boldsymbol{x}=(x_1,\cdots,x_n)$ 为元件的状态向量. 假定系

统也只有正常与失效状态，其状态完全由元件的状态决定. 用$\varphi(\boldsymbol{x})$记系统状态. 再记$(0_i,\boldsymbol{x})$，$(1_i,\boldsymbol{x})$分别为第i个分量为0或1的状态向量，$(\cdot_i,\boldsymbol{x})$为第i个分量的任意值.

定义 33.5.1 $\varphi(\boldsymbol{x})$称作一个n阶**单调关联系统**(monotone coherent system)，若

(1) $\boldsymbol{x}\leqslant\boldsymbol{y}$，则$\varphi(\boldsymbol{x})\leqslant\varphi(\boldsymbol{y})$.

(2) $\forall\ i$，$\exists\ (\cdot_i,\boldsymbol{x})$使$\varphi(0_i,\boldsymbol{x})<\varphi(1_i,\boldsymbol{x})$

这里$\boldsymbol{x}\leqslant\boldsymbol{y}$是指$x_i\leqslant y_i$，$i=1,\cdots,n$.

例 33.5.2 串联系统 $\varphi(\boldsymbol{x}) = \min\limits_{1\leqslant i\leqslant n} x_i = \prod\limits_{i=1}^{n} x_i$.

并联系统 $\varphi(\boldsymbol{x}) = \max\limits_{1\leqslant i\leqslant n} x_i = 1 - \prod\limits_{i=1}^{n}(1 - x_i)$.

k-out-of-n(F)系统：系统由n个元件组成，当k或k个以上元件失效时系统失效.

$$\varphi(x) = \begin{cases} 1, & \text{若}\sum\limits_{i=1}^{n} x_i > n - k, \\ 0, & \text{其它}. \end{cases}$$

若元件i的寿命记作x_i，用$x_i(t)$记时刻t元件i的状态，$\boldsymbol{X}(t)=(x_1(t),\cdots,x_n(t))$. 系统在时刻$t$的状态用$\varphi(\boldsymbol{X}(t))$记. 则系统在$t$正常的概率为

$$R(t) = P\{\varphi(\boldsymbol{X}(t)) = 1\} = E\varphi(\boldsymbol{X}(t)).$$

33.6 网络系统可靠度计算

网络系统是由一些节点及弧连接而成，其中弧表示系统中的元件，节点表示元件之间的关系. 假定节点完全可靠，而系统及弧只有正常、失效两种状态，分别用1,0表示. 进一步设弧之间独立.

网络可靠度问题可形象地描述为：设想在一个无向网络中从

输入节点 S 发出一个信息，通过网络中不完全可靠的弧的传送，问输出节点 t 接收到该信息的概率(即网络可靠度)有多大？

现设网络 S 由 n 条弧组成，用 x_i 记第 i 条弧的状态. $\boldsymbol{x}=(x_1,\cdots,x_n)$为弧的状态向量，用 $\varphi(\boldsymbol{x})$表示网络的状态.

定义 33.6.1 $\boldsymbol{x}$ 称作 S 的一条**路向量**(path vector)，若 $\varphi(\boldsymbol{x})=1$. 进一步，$\forall\ \boldsymbol{y}<\boldsymbol{x}$，$\varphi(\boldsymbol{y})=0$，则称 $\boldsymbol{x}$ 是**最小路向量**(minimal path vector)，这里 $\boldsymbol{x}<\boldsymbol{y}$ 表示分量间至少有一个严格不等号成立. 相应的 $c_1(\boldsymbol{x})=\{j:x_j=1\}$称作**最小路集**(minimal path set)，简称**最小路**. 最小路中的元素个数称为其长度.

$\boldsymbol{x}$ 称作 S 的一个**割向量**(cut vector)，若 $\varphi(\boldsymbol{x})=0$. 进一步，$\forall\ \boldsymbol{y}>\boldsymbol{x}$，$\varphi(\boldsymbol{y})=1$，则称 $\boldsymbol{x}$ 为**最小割向量**(minimal cut vector). 相应的 $c_0(\boldsymbol{x})=\{j:x_j=0\}$称作**最小割集**(minimal cut set)，简称**最小割**.

按定义最小路是一些弧的下标集，同时又指与这些下标相应的弧的集合. 若记 S 的所有最小路为 $A_1,\cdots,A_m$. 仍用 A_i 表示第 i 条最小路中的弧都正常这一事件，于是

$$\{系统正常\}=\{\varphi(\boldsymbol{x})=1\}=\bigcup_{i=1}^{m}A_i.$$

系统的可靠度为 $R=P\{\varphi(\boldsymbol{x})=1\}=P\left(\bigcup_{i=1}^{m}A_i\right)$.

相仿，若 S 的所有最小割为 $B_1,\cdots,B_l$，则$\{系统失效\}=\{\varphi(\boldsymbol{x})=0\}=\bigcup_{i=1}^{l}B_i$，$R=1-P\left(\bigcup_{i=1}^{l}B_i\right)$.

因此，网络可靠度的计算归结为

(1) 求出所有最小路(割)$A_1,\cdots,A_m$；

(2) 求事件并的概率 $P\left(\bigcup_{i=1}^{m}A_i\right)$.

1. 求所有最小路的方法

定义 33.6.2 给定 l 个节点的网络，定义相应的**邻接矩阵**(connection matrix)$\boldsymbol{C}=(c_{ij})$，其中

$$c_{ij}=\begin{cases}0,\text{若节点 } i,j \text{ 间无弧直接相连},\\ x,\text{若节点 } i \text{ 经过弧 } x \text{ 可直接到达 } j.\end{cases}$$

记 $c_{ij}^{(2)}$ 为节点 i 到 j 的长度为 2 的所有可能的最小路. $c_{ij}^{(2)}$ 可通过矩阵"乘法"公式得到:

$$c_{ij}^{(2)} = \sum_{k=1}^{l} c_{ik} \cdot c_{kj}, \quad i,j = 1,\cdots,l.$$

这里 $\sum$ 表示集合的并, $\cdot$ 表示交(通常省略). 并且约定右端要利用集合运算规则进行归并,且只保留长度为 2 的最小路.

对 $c_{ij}^{(r)}$ 可相仿定义并依 r 递归求得. 由于最小路长度最多为 $l-1$,因此若记输入、输出结点分别为 1,2,则结点 1,2 间的所有最小路为 $c_{12}, c_{12}^{(2)}, \cdots, c_{12}^{(l-1)}$.

例 33.6.3 桥形网络 一桥形网络图,如图 33.1

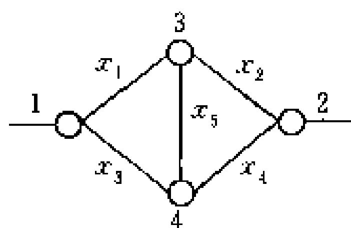


图 33.1

定义 33.6.4 邻接矩阵

$$\boldsymbol{C}=\begin{pmatrix}0 & 0 & x_1 & x_3\\ 0 & 0 & x_2 & x_4\\ x_1 & x_2 & 0 & x_5\\ x_3 & x_4 & x_5 & 0\end{pmatrix},$$

容易得到 $c_{12}=0, c_{11}^{(2)}=0, c_{12}^{(2)}=x_1x_2+x_3x_4, c_{13}^{(2)}=x_3x_5, c_{14}^{(2)}=x_1x_5$, $c_{12}^{(3)}=x_2x_3x_5+x_1x_4x_5$. 梯形网络中所有最小路为 x_1x_2, x_3x_4, $x_1x_4x_5, x_2x_3x_5$.

2. 求可靠度的算法

问题归结为求事件并的概率 $R=P\left(\bigcup_{i=1}^{m} A_i\right)$.

(1) 利用初等概率公式

$$P\left(\bigcup_{i=1}^{m}A_i\right)=\sum_{i=1}^{m}(-1)^{i-1}S_i,\quad S_i=\sum_{1\leqslant j_1<\cdots<j_i\leqslant m}P(A_{j_1}\cdots A_{j_i}).$$

当 m 大时方法不实用.

(2) 不交化算法求可靠度 R

设网络共有 n 个结点，所有最小路 $A_1,\cdots,A_m$ 已求得. 不交化可分两步：

(i) 取出长度为 $n-1$ 的所有最小路 $A_{k+1},\cdots,A_m$. $A_j(j=k+1,\cdots,m)$中未出现的弧 x 用 $\bar{x}$ 补上. 所得的这些事件互相不交，且与长度小于 $n-1$ 的所有最小路表示的事件不交.

(ii) 对长度小于 $n-1$ 的最小路 $A_1,\cdots,A_k$ 由如下手续不交：$A_1+A_2\overline{A}_1+A_3\overline{A}_2\overline{A}_1+\cdots+A_k\overline{A}_{k-1}\cdots\overline{A}_1$. 最后把 a,b 中结果合并即得不交和.

例 33.6.5 图 33.2 中网络输入、输出结点 1,2 之间的所有最小路为

$$fg,abc,adg,fec,abeg,adec,fdbc.$$

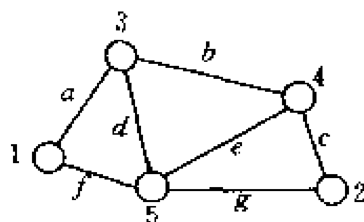


图 33.2

由于网络结点数为 5，故长度为 4 的所有最小路不交化后为

$$ab\bar{c}\bar{d}e\bar{f}g+a\bar{b}cde\bar{f}\bar{g}+\bar{a}bcd\bar{e}f\bar{g}.$$

下面对 $fg+abc+adg+fec$ 化成不交和(为符号简单上式中用+代替∪).

(1) 取出 fg，保留.

(2) 取出 abc，与前面的保留项不交. 因此，这一步保留.

$$abc\,\overline{fg}=abc(\bar f+\bar g)=abc(\bar f+f\bar g).$$

注意在这一步中$\bar f+\bar g$本身也要进行不交化，同时利用如下运算法则进行简化.

$$x+\bar x=1,\ x\bar x=0,\ \overline{x+y}=\bar x\bar y,\ \overline{xy}=\bar x+\bar y,\ x+xy=x.$$

(3) 取出adg，与前面的所有保留项不交，故保留

$$adg\,\bar f\,\overline{bc}=adg\,\bar f(\bar b+b\bar c).$$

(4) 取出fec，与前面的保留项不交.

$$fec\bar g\,\overline{ab}=fec\bar g(\bar a+a\bar b).$$

注意，此时不必与(3)中的保留项不交化，因为上述项中有$\bar g$，而(3)中保留项有g.

因此，最终的不交和为

$$fg+abc(\bar f+f\bar g)+adg\bar f(\bar b+bc)+fec\bar g(\bar a+a\bar b)+ab\bar c\bar d e\bar f g+a\bar b c d e\bar f\,\bar g+\bar a b c d\bar e f g.$$

对大型网络求所有最小路及不交化手续可以利用计算机来实现.

网络可靠度目前研究的问题有：简化网络的方法；k个节点相通的可靠度问题；算法复杂性研究等.

33.7 故障树分析

故障树分析(fault tree analysis，FTA)是20世纪60年代发展起来的一种可靠性分析方法.用于定量评定不希望的事件(称作顶事件)T发生的概率，并找出导致T发生的各种基本事件的所有可能的组合.由于FTA可用一棵倒长的树来表示，故称故障树.

方法 33.7.1 FTA的程序

(1) 选择顶事件T 顶事件的选择必须有明确定义，且能进一步分解.

（2）建立故障树　由顶事件 T 按逻辑关系逐级分解来建立树. 若 T 的发生由次级事件 A,B 中任一发生而引起，则由"或门"把 T 与 A,B 相联系；若 A,B 同时发生 T 才发生，则由"与门"相联系. 对 A,B 又可分析下去，一直到不能分解的基本事件为止.

故障树的建立需要专业人员与可靠性工作者结合.

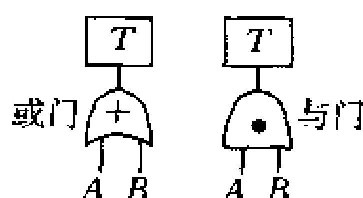


图　33.3

（3）故障树的评定　假定故障树由 n 个基本失效事件组成. 用 $x_i=1$ 或 0 表示基本事件 i 出现或不出现，$\boldsymbol{x}=(x_1,\cdots,x_n)$. 顶事件 T 的状态由 $\varphi(\boldsymbol{x})$ 表示，$\varphi(\boldsymbol{x})=1$ 或 0 时表示 T 发生或不发生.

定义 33.7.2　若 $\varphi(\boldsymbol{x})=1$，称 $\boldsymbol{x}$ 为 φ 的一个**割向量**；若又有 $\forall\ \boldsymbol{y}<\boldsymbol{x},\varphi(\boldsymbol{y})=0$，则 $\boldsymbol{x}$ 是一个最小割向量. 记 $c_1(\boldsymbol{x})=\{j:x_j=1\}$，称为相应于 $\boldsymbol{x}$ 的**最小割集**，简称**最小割**.

若记 $k_1,\cdots,k_l$ 为 φ 的所有最小割. 仍用 k_i 表示该最小割相应的基本事件发生的交. 因此，$\{T\text{ 发生}\}=\{\boldsymbol{x}:\varphi(\boldsymbol{x})=1\}=\bigcup_{i=1}^{l}k_i$. 于是，故障树的评定要求出所有最小割 $k_1,\cdots,k_l$ 及概率 $P\left(\bigcup_{i=1}^{l}k_i\right)$.

方法 33.7.3　下行法求所有最小割　由顶事件 T 往下列表逐级进行. 若 T 下面一级是"与门"，则把该门下所有输入都排在同一行中；若 T 下面是"或门"，则把每个输入事件单列一行. 然后依次做下去，一直到基本事件为止. 最后进行检查，保留两两互不包含的割集，即为所有最小割.

例 33.7.4　求图 33.4 所有最小割.

列表如下：

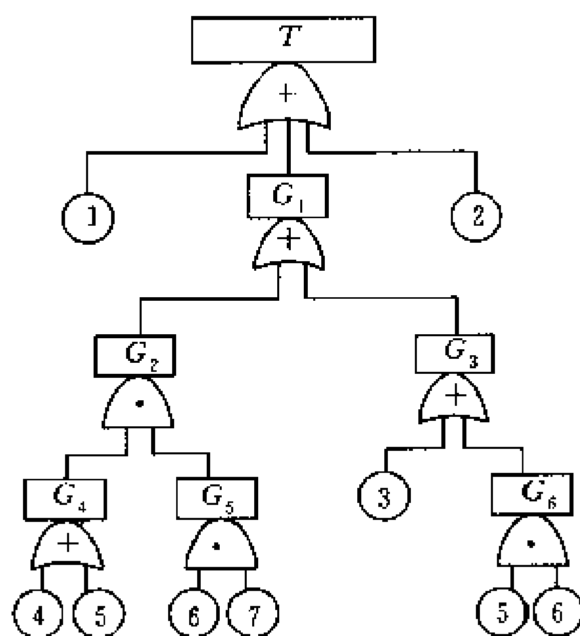


图 33.4

步骤	1	2	3	4	5
	1	1	1	1	1
	2	2	2	2	2
	G_1	G_2	G_4,G_5	4,G_5	4,6,7
		G_3	3	5,G_5	5,6,7
			G_6	3	3
				5,6	5,6

由上表第5列得6个割集：

{1},{2},{3},{5,6},{4,6,7},{5,6,7}.

但由于{5,6}⊂{5,6,7},故应删去{5,6,7}.留下的无相互包含的关系.故所有最小割为{1},{2},{3},{5,6},{4,6,7}.

在已知基本事件发生的概率及互相独立的条件下，$P\left(\bigcup_{i=1}^{l} k_i\right)$ 的求法与网络可靠度求法相同.

FTA目前研究的问题有：对大型问题有效的算法与软件；事件的逻辑关系中出现"非门"等更一般的事件树的分析方法；计算

机辅助 FTA 等.

33.8 可修复系统可靠性分析

33.8.1 可修复系统的可靠性指标

对可修复系统的状态，它总是正常与失效相交替的一个过程. 其可靠性指标有：系统首次失效时间 T 的分布；**可用度**(availability) $A(t)=P\{$系统在时刻 t 正常$\}$；以及 $(0,t]$ 中系统失效次数 $N(t)$. 此外，若考虑到修理，则有修理工在 t 时忙的概率 $B(t)$. 若 $t\to\infty$ 时，$A(t)$，$B(t)$ 或 $\frac{1}{t}EN(t)$ 的极限存在，则分别称其为稳态下的量.

33.8.2 单部件可修复系统

设系统由单部件组成，$t=0$ 时开始运行，工作与修理相交替，相继工作时间为非负随机变量 $X_1,X_2,\cdots$，独立同分布 $F(t)$，相继修复时间为 $Y_1,Y_2,\cdots$，独立同分布 $G(t)$. 并假定 $\{X_n\}$ 与 $\{Y_n\}$ 独立，及 $EX_1=\frac{1}{\lambda}$，$EY_1=\frac{1}{\mu}$，$0<\lambda<\infty$，$0<\mu\leqslant\infty$.

系统首次失效时间为 X_1，有可靠度函数 $R(t)=\overline{F}(t)$，$t\geqslant 0$. $EX_1=\frac{1}{\lambda}$.

系统可用度 $A(t)$ 相应的 LS 变换为

$$\hat{a}(s)=\frac{1-\hat{f}(s)}{1-\hat{f}(s)\hat{g}(s)},\quad s>0. \tag{33.10}$$

这里 $\hat{a}$ 表示 A 的 LS 变换 $\hat{a}(s)=\int_0^\infty e^{-st}dA(t)$，$s>0$. 稳态可用度 $A=\mu/(\lambda+\mu)$.

$(0,t]$ 中系统平均失效数 $M(t)$ 相应的 LS 变换为

$$\hat{m}(s) = \frac{\hat{f}(s)}{1 - \hat{f}(s)\hat{g}(s)}, \quad s > 0. \qquad (33.11)$$

稳态下的平均失效率 $\nu = \lim_{t\to\infty} \frac{1}{t} M(t) = \frac{\lambda\mu}{\lambda+\mu}$.

特别,若 X_1, Y_1 分别有参数为 λ, μ 的指数分布时

$$A(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} \mathrm{e}^{-(\lambda+\mu)t}, \quad t \geqslant 0. \qquad (33.12)$$

$$M(t) = \frac{\lambda\mu}{\lambda+\mu} t + \left(\frac{\lambda}{\lambda+\mu}\right)^2 (1 - \mathrm{e}^{-(\lambda+\mu)t}), \quad t \geqslant 0.$$

33.8.3 马尔可夫可修复系统

所谓马尔可夫可修复系统是指系统中的部件互相独立,其寿命及其修复时间均为指数分布的系统.此时系统的状态可以由一个时齐的马尔可夫过程来描述.

设$\{X(t), t \geqslant 0\}$是有限状态空间 $J = \{1, \cdots, m+n\}$上的一个时齐马尔可夫过程.相应的转移率矩阵为 $\boldsymbol{Q} = (q_{ij})$,

$$q_{ii} \leqslant 0,\ q_{ij} \geqslant 0 (i \neq j),\ \sum_{j\in J} q_{ij} = 0, \quad i, j \in J. \qquad (33.13)$$

记状态概率(行向量)$\boldsymbol{p}(t)$,其第 j 个分量 $p_j(t) = P\{X(t) = j\}, j \in J$.初始状态概率 $\boldsymbol{\alpha} = \boldsymbol{p}(0)$. $\boldsymbol{p}(t)$可表为

$$\boldsymbol{p}(t) = \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{Q}t}, \quad t \geqslant 0 \qquad (33.14)$$

或 $\boldsymbol{p}(t)$相应的 L 变换

$$\boldsymbol{p}^*(s) = \int_0^\infty \boldsymbol{p}(t) \mathrm{e}^{-st} \mathrm{d}t = \boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{Q})^{-1},\ \boldsymbol{I}\ \text{单位阵} \qquad (33.15)$$

其稳态概率 $\boldsymbol{p} = \lim_{t\to\infty} \boldsymbol{p}(t)$由

$$\boldsymbol{p}\boldsymbol{Q} = 0, \quad \boldsymbol{p}\boldsymbol{1} = 1 \qquad (33.16)$$

解出,其中 $\boldsymbol{1}$ 是分量全为 1 的列向量.

假定从系统可靠性角度来讲,系统的状态可分为两类:$W = \{1, \cdots, m\}$为正常状态,$B = \{m+1, \cdots, m+n\}$为失效状态.系统的可靠性指标为

（1）可用度 $A(t)$ 及稳态可用度 A

$$A(t)=\sum_{j\in W}p_j(t),\ t\geqslant 0.\quad A=\sum_{j\in W}p_j.$$

（2）系统可靠度函数 $R(t)$ 及首次失效时间 T 的均值

$$R(t)=P(T>t)=\boldsymbol{\alpha}_W e^{\boldsymbol{Q}_W t},\ t\geqslant 0.\quad ET=-\boldsymbol{\alpha}_W\boldsymbol{Q}_W^{-1}\mathbf{1}.$$

其中 $\boldsymbol{\alpha}_W=(\alpha_j,j\in W)$，$\boldsymbol{Q}_W=(q_{ij},i,j\in W)$.

（3）稳态下平均开工时间 MUT(mean up-time)，平均停工时间 MDT (mean down-time) 及平均周期长 MCT (mean cycle time).

公式 33.8.1 $\mathrm{MUT}=-\dfrac{\boldsymbol{p}_W\mathbf{1}}{\boldsymbol{p}_W\boldsymbol{Q}_W\mathbf{1}}=\dfrac{\boldsymbol{p}_W\mathbf{1}}{\boldsymbol{p}_W\boldsymbol{Q}_{WB}\mathbf{1}}$，$\mathrm{MDT}=-\dfrac{\boldsymbol{p}_B\mathbf{1}}{\boldsymbol{p}_W\boldsymbol{Q}_W\mathbf{1}}=\dfrac{\boldsymbol{p}_B\mathbf{1}}{\boldsymbol{p}_W\boldsymbol{Q}_{WB}\mathbf{1}}$，$\mathrm{MCT}=(\boldsymbol{p}_W\boldsymbol{Q}_{WB}\mathbf{1})^{-1}$，这里 $\boldsymbol{p}_B=(p_j,j\in B)$，$\boldsymbol{Q}_{WB}=(q_{ij},i\in W,j\in B)$.

（4）稳态下系统平均失效率 M

公式 33.8.2 $M=\boldsymbol{p}_W\boldsymbol{Q}_{WB}\mathbf{1}=(\mathrm{MCT})^{-1}$.

（5）瞬时及稳态失效频率 $\nu(t)$ 及 ν：

公式 33.8.3 $\nu(t)=-\boldsymbol{p}_W(t)\boldsymbol{Q}_W\mathbf{1}=\boldsymbol{p}_W(t)\boldsymbol{Q}_{WB}\mathbf{1},t\geqslant 0.\ \nu=M=\boldsymbol{p}_W\boldsymbol{Q}_{WB}\mathbf{1}$.

例 33.8.4 2-out-of-3(F)系统　设系统由独立同型部件组成，部件的寿命及修复时间分别为 λ,μ 的指数分布. 系统中有一个修理工. $t=0$ 时系统开始运行，所有部件正常.

定义状态 i 为系统中恰有 i 个部件失效，$i=0,1,2$. 状态空间

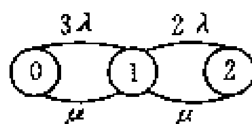


图　33.5

$J=W\cup B$，$W=\{0,1\}$，$B=\{2\}$. 状态间的转移率由图 33.5 标出初始状态概率为 $\boldsymbol{\alpha}=(1,0,0)$，转移率矩阵为

$$Q=\begin{pmatrix}-3\lambda & 3\lambda & 0\\ \mu & -(2\lambda+\mu) & 2\lambda\\ 0 & \mu & -\mu\end{pmatrix}.$$

为表达式简单起见只列出稳态或平均结果，记号同前.

$$p_0=(6\rho^2+3\rho+1)^{-1},\quad p_1=3\rho p_0,\quad p_2=6\rho^2 p_0,\quad \rho=\frac{\lambda}{\mu}.$$

$$ET=\frac{5}{6}\frac{1}{\lambda}+\frac{\mu}{6\lambda^2}.$$

$$A=p_0+p_1=(1+3\rho)p_0.$$

$$\nu=M=6\lambda\rho p_0.$$

$$\text{MUT}=\frac{1+3\rho}{6\lambda\rho},\quad \text{MDT}=\frac{1}{\mu},\quad \text{MCT}=(6\lambda\rho p_0)^{-1}.$$

33.8.4 补充变量法

对可修复系统进行分析时，若寿命或修复时间不是指数分布，则马尔可夫性遭到破坏. 此时若能通过引入新的补充变量，使系统的状态用一个高维的马尔可夫过程来描述，则可建立状态概率满足的偏微分方程，于是通过对这些方程的求解获得系统性状的定量评定. 因此，补充变量法是分析复杂系统可靠性时的一个重要工具.

由如下模型来说明补充变量方法. 设系统由两个同型部件及一个修理设备组成. 记 $N(t)$ 为时刻 t 失效的部件数(0，1 或 2). 设系统中部件的失效规律为

$$P(N(t+\Delta t)=1\,|\,N(t)=0)=\lambda_0\Delta t+o(\Delta t),$$

$$P(N(t+\Delta t)=2\,|\,N(t)=1)=\lambda_1\Delta t+o(\Delta t).$$

失效部件按顺序进行修理，且修复如新，修复后立即投入工作. 进一步设修理与运行这些过程都是互相独立的. 修理时间有一般连续分布 $G(t)$，密度为 $g(t)$，均值 $\frac{1}{\mu}$ 有限. 系统失效当且仅当两个部件都处在失效状态时.

由于修理时间遵从一般分布，故 $N(t)$ 对任一时刻不具有马尔可夫性，因此，$\{N(t),t\geqslant 0\}$ 不是马尔可夫过程. 例如已知 $N(t)=1$，此时仍不足以确定 t 以后 $N(t)$ 的概率规律. 因为在时刻 t 已失效部件所得到的修理时间长短显然影响到 $N(t)$ 的变化规律.

为此引进一个补充变量 $X(t)$，记录时刻 t 正在修理的部件所占用的修理时间. 约定当 $N(t)=0$ 时，$X(t)=0$. 于是 $\{N(t),X(t);t\geqslant 0\}$ 是一个状态空间 $J\times[0,\infty)$ 上的马尔可夫过程，这里 $J=\{0,1,2\}$. 这个过程的马尔可夫性直观上容易得到证实：对任意时刻 t，若 $N(t)$，$X(t)$ 给定，则过程在 t 以后的概率规律与 t 以前的历史无关.

1. 系统可用度

对任意 $t\geqslant 0,x\geqslant 0$，令

$$p_0(t)=P(N(t)=0),$$

$$p_1(t,x)\mathrm{d}x=P(N(t)=1,x<X(t)\leqslant x+\mathrm{d}x),$$

$$p_2(t,x)\mathrm{d}x=P(N(t)=2,x<X(t)\leqslant x+\mathrm{d}x).$$

初始条件 $p_0(0)=1,p_1(0,x)=p_2(0,x)=0$.

由上述诸式可以导出如下方程

$$\frac{\mathrm{d}}{\mathrm{d}t}p_0(t)=-\lambda_0 p_0(t)+\int_0^\infty p_1(t,x)\mu(x)\mathrm{d}x, \tag{33.17}$$

$$\frac{\partial}{\partial t}p_1(t,x)+\frac{\partial}{\partial t}p_1(t,x)=-[\lambda_1+\mu(x)]p_1(t,x),\quad x>0, \tag{33.18}$$

$$\frac{\partial}{\partial t}p_2(t,x)+\frac{\partial}{\partial x}p_2(t,x)=-\mu(x)p_2(t,x)+\lambda_1 p_1(t,x),\quad x>0, \tag{33.19}$$

$$p_1(t,0)=\lambda_0 p_0(t)+\int_0^\infty p_2(t,x)\mu(x)\mathrm{d}x,$$

$$p_2(t,0)=0,$$

$$p_0(t)+\int_0^\infty p_1(t,x)\mathrm{d}x+\int_0^\infty p_2(t,x)\mathrm{d}x=1.$$

这里 $\mu(x)=g(x)/\overline{G}(x)$ ($\overline{G}=1-G$),是 G 相应的修复率函数.(33.17)可如下推出:为使 $N(t+\Delta t)=0$,必须 $N(t)=0$,且$(t, t+\Delta t]$内无故障发生;或 $N(t)=1, X(t)=x>0$,且在$(t,t+\Delta t]$中失效部件修复,这里 x 可以是任意正实数,于是

$$p_0(t+\Delta t)=p_0(t)(1-\lambda_0\Delta t)+\int_0^{\infty}p_1(t,x)\mu(x)\mathrm{d}x\Delta t+o(\Delta t).$$

经过简化,即可得(33.17).另外(33.18),(33.19)可相仿获得.为求出可用度,可对偏微分方程组两端做 L 变换(对变量 t).

利用初始条件可得关于 x 的微分积分方程

$$sp_0^*(s)-1=-\lambda_0p_0^*(s)+\int_0^{\infty}p_1^*(s,x)\mu(x)\mathrm{d}x,$$

$$sp_1^*(s,x)+\frac{\mathrm{d}}{\mathrm{d}x}p_1^*(s,x)=-(\lambda_1+\mu(x))p_1^*(s,x),\quad x>0,$$

$$sp_2^*(s,x)+\frac{\mathrm{d}}{\mathrm{d}x}p_2^*(s,x)=-\mu(x)p_2^*(s,x)+\lambda_1p_1^*(s,x),\ x>0.$$

$$p_1^*(s,0)=\lambda_0p_0^*(s)+\int_0^{\infty}p_2^*(s,x)\mu(x)\mathrm{d}x,$$

$$p_2^*(s,0)=0,$$

$$p_0^*(s)+\int_0^{\infty}p_1^*(s,x)\mathrm{d}x+\int_0^{\infty}p_2^*(s,x)\mathrm{d}x=\frac{1}{s}.$$

这里用 $p_0^*(\cdot)$等表示 $p_0(\cdot)$的 L 变换,解之可得

$$p_0^*(s)=[1-g^*(s)+g^*(s+\lambda_1)]/\alpha(s),$$

$$p_1^*(s,x)=\lambda_0\overline{G}(x)\mathrm{e}^{-(s+\lambda_1)x}/\alpha(s),$$

$$p_2^*(s,x)=\lambda_0\overline{G}(x)\mathrm{e}^{-sx}(1-\mathrm{e}^{-\lambda_1x})/\alpha(s),$$

其中 $\quad \alpha(s)=sg^*(s+\lambda_1)+(s+\lambda_0)[1-g^*(s)]$.

系统的瞬时可用度为

$$A(t)=p_0(t)+\int_0^{\infty}p_1(t,x)\mathrm{d}x,$$

相应的 L 变换为

$$A^*(s)=(s+\lambda_1)[1-g^*(s)+g^*(s+\lambda_1)] + \lambda_0[1-g^*(s+\lambda_1)]/[(s+\lambda_1)\alpha(s)].$$

稳态下可用度 $A=\lim\limits_{t\to\infty}A(t)=\lim\limits_{s\to0^+}sA^*(s)$. 即

$$A=[\lambda_0\mu+\mu(\lambda_1-\lambda_0)g^*(\lambda_1)]/[\lambda_0\lambda_1+\lambda_1\mu g^*(\lambda_1)].$$

2. 系统可靠度

为求可靠度 $R(t)$,只要把原过程 $\{N(t),X(t),t\geqslant0\}$ 中的状态 $\{2\}\times(0,\infty)$ 当作吸收态,这样定义了一个新的马氏过程 $\{\tilde{N}(t),\tilde{X}(t),t\geqslant0\}$ 记

$$q_0(t)=P(\tilde{N}(t)=0),$$

$$q_1(t,x)\mathrm{d}x=P(\tilde{N}(t)=1,x<\tilde{X}(t)\leqslant x+\mathrm{d}x).$$

初条件 $q_0(0)=1,q_1(0,x)=0$.

于是可靠度可表为

公式 33.8.5 $R(t)=q_0(t)+\int_0^\infty q_1(t,x)\mathrm{d}x.$

通过导出 $q_0(t),q_1(t,x)$ 满足的偏微分方程,由 L 变换的方法可得 $q_0(t),q_1(t,x)$ 相应的 L 变换式 $q_0^*(s),q_1^*(s,x)$.

$$q_0^*(s)=\{s+\lambda_0[1-g^*(s+\lambda_1)]\}^{-1},$$

$$q_1^*(s,x)=\lambda_0\bar{G}(x)\mathrm{e}^{-(s+\lambda_1)x}q_0^*(s).$$

因此 $R(t)$ 相应的 L 变换为

公式 33.8.6

$$R^*(s)=\{s+\lambda_1+\lambda_0[1-g^*(s+\lambda_1)]\}q_0^*(s)/(s+\lambda_1).$$

系统首次失效前平均时间 T 为

公式 33.8.7

$$T=\int_0^\infty R(t)\mathrm{d}t=\lim_{s\to0^+}R^*(s)=\frac{1}{\lambda_1}+\frac{1}{\lambda_0[1-g^*(\lambda_1)]}.$$

从上面的模型来看,使用补充变量法的步骤为:引入适当的补充变量,使描述系统状态的过程成为一个马尔可夫过程;建立状态概率满足的偏微分方程;由 L 变换解方程;求稳态解或平均值. 由于补充变量的个数不能太多,否则实际求解常常变得不可能,因此限制了其应用的范围.

分析更复杂的可修复系统需要进一步的工具，例如马尔可夫更新过程等，请参阅第 27 章及文献[47]，[97]，[101]，[102]等.

33.9 更换策略

系统在运行中失效一般比失效前更换带来的损失更大. 因此，预防性**更换策略**(replacement policy)的研究有实际意义.

设系统的寿命分布为 $F(t)$. 失效前或后每更换一次的费用分别为 c_p, c_f. 假定更换是瞬时的. 记 $c(t)$ 为 $(0,t]$ 中更换的总费用. $c=\lim_{t\to\infty}\frac{1}{t}Ec(t)$，为长期运行下单位时间中的期望费用. 问题为研究在不同的更换方式下使 c 达最小的更换策略.

方法 33.9.1 按年龄更换策略(age policy)是指到一定的年龄 T(若系统此时未失效)或在其失效时进行更换. 此时

$$c = c(T) = \frac{c_f F(T) + c_p \overline{F}(T)}{\int_0^T \overline{F}(t)\mathrm{d}t}.$$

最佳更换年龄 T^* 由使上式达最小而求得.

方法 33.9.2 成批更换策略(block policy)是指在周期时刻 $kT(k=1,2,\cdots)$ 上或系统失效时进行更换. 此时

$$c = c(T) = \frac{1}{T}[c_p + c_f EN(T)],$$

其中 $N(t)$ 为 $(0,t]$ 中失效次数. 使上式达最小求出最佳更换周期 T^*.

方法 33.9.3 小修策略(minimal repair policy)是指系统在 kT 周期性地进行预防性更换($k=1,2,\cdots$)，若在周期中失效则进行小修，使系统恢复到失效前的状态(即小修后系统的失效率 $r(t)$ 不变). 若每次小修的费用为 c_m，则

$$c = c(T) = \frac{1}{T}\left[c_p + c_m\int_0^T r(u)\mathrm{d}u\right].$$

使上式达最小求出最佳更换周期.

更换策略研究中更进一步的问题有考虑折扣因子,视情况更换或维修,以及费用依赖于修理次数或年龄等模型.

33.10 寿命数据分析

寿命数据统计分析是可靠性工作的基础.这种分析有其自身的特点:数据的不完全性.原因可能是现场数据不全;或对长寿命产品特别是元器件,寿命试验不可能进行到所有受试件全部失效,因而不能获得所有样品的寿命数据.此外,加速寿命试验数据,可修复系统相邻故障间隔的数据,以及由部件寿命试验数据如何获得系统可靠性指标等,都需要不同的处理方法.

寿命试验中最常用的有:

(1) 定时截尾方案 n 个独立同分布的样品 $X_1,\cdots,X_n$ 从 $t=0$开始受试,试验到事先规定的时间 t_0 终止.试验中的失效件不进行替换.假定试验结束时获得 r 个失效数据

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(r)} \leqslant t_0,$$

这里 $x_{(1)},\cdots,x_{(r)}$是 $X_1,\cdots,X_n$ 从小到大的顺序统计量的头 r 个观察值.注意,r 是一个随机变量,可能值为 $0,1,\cdots,n$.试验结束时的失效个数是其观察值.

(2) 定数截尾方案 与定时方案相仿,试验进行到第 r 个(事先规定的)失效发生时终止,可获数据

$$x_{(1)} \leqslant \cdots \leqslant x_{(r)},$$

这里试验终止的时间 $X_{(r)}$是随机的.

其它试验方式还有定时与定数截尾相结合的混合方案或序贯方案等.

33.10.1 指数总体的情形

设总体 X 有指数分布 $F(t)=1-\mathrm{e}^{-\frac{t}{\theta}}, t\geqslant 0$. 平均寿命 $EX=\theta$ 为未知参数.

1. 定数截尾方案

记 $T(r)=\sum_{i=1}^{r}X_{(i)}+(n-r)X_{(r)}$. 则 θ 的极大似然估计(MLE)为 $\hat{\theta}=\frac{1}{r}T(r)$. θ 的置信度为 $1-\alpha$ 的双侧置信区间为 $\left[\frac{2nT(r)}{\chi_{2r}^2\left(1-\frac{\alpha}{2}\right)},\frac{2nT(r)}{\chi_{2r}^2\left(\frac{\alpha}{2}\right)}\right]$, 这里 $\chi_{2r}^2(\beta)$ 为 χ_{2r}^2 分布的 β 分位点, 即 $P\{\chi_{2r}^2\leqslant\chi_{2r}^2(\beta)\}=\beta$.

2. 定时截尾方案

记 $T(t_0)=\sum_{i=0}^{r}X_{(i)}+(n-r)t_0$, 则 $\hat{\theta}=\frac{1}{r}T(t_0)$. θ 的置信度至少为 $1-\alpha$ 的双侧置信区间为 $\left[-\frac{t_0}{\log(1-a)},-\frac{t_0}{\log(1-b)}\right]$, 这里

$$a=\left[1+\frac{n-r+1}{r}F_{2(n-r+1),2r}\left(1-\frac{\alpha}{2}\right)\right]^{-1},$$

$$b=\left[1+\frac{n-r}{r+1}F_{2(n-r),2(r+1)}\left(\frac{\alpha}{2}\right)\right]^{-1}.$$

而 $F_{m,n}(\alpha)$ 为 $F_{m,n}$ 分布的 α 分位点.

33.10.2 韦布尔总体的情形

设总体 X 有韦布尔(Weibull)分布 $F(t)=1-\mathrm{e}^{-\frac{t^m}{T_0}}, t>0, T_0, m$ 为未知参数. 在定时或定数截尾方案下, 数据统一记为 $x_{(1)}\leqslant\cdots\leqslant x_{(r)}\leqslant x_{(s)}$, 这里

$$x_{(s)}=\begin{cases}T_0, & \text{定时方案下},\\ x_{(r)}, & \text{定数方案下}.\end{cases}$$

未知参数 m, T_0 的 MLE 由下面的方程迭代解出

$$\frac{1}{m}=\frac{\sum_{i=1}^{r} x_i^m \log x_i+(n-r) x_r^m \log x_r}{\sum_{i=1}^{r} x_i^m+(n-r) x_r^m}-\frac{1}{r} \sum_{i=1}^{r} \log x_i,$$

$$T_0=\frac{1}{r}\left[\sum_{i=1}^{r} x_i^m+(n-r) x_r^m\right].$$

上述两式中为记号简单起见，用 x_i 来代替 $x_{(i)}$.

33.10.3 Γ分布总体的情形

设总体 X 有分布密度 $f(t)=\frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \mathrm{e}^{-\lambda t}, t>0, \lambda, \alpha>0$ 为未知参数. 在完全样本 $x_1, \cdots, x_n$ 下，λ, α 的矩法估计量为

$$\tilde{\lambda}=(n-1) \bar{x} / \sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2, \quad \tilde{\alpha}=\tilde{\lambda} \bar{x},$$

其中 $\bar{x}=\frac{1}{n} \sum_{i=1}^{n} x_i$.

33.11 软件可靠性

70 年代起，计算机已渗透到现代社会生活中的各个方面. 宇航，空中交通控制系统，通讯，银行系统，防务系统等都要求计算机有极高的可靠性. 因此，开发、运行和维护与复杂的系统功能相适应的软件，以及软件本身的可靠性就越来越受到重视.

所谓软件失效或故障是指软件不能完成预定的功能或达不到设计规范所规定的标准(如精度或响应时间要求).

定义 33.11.1 在特定的环境中，在一定的时间间隔中，软件正常运行的概率称为**软件可靠度**(software reliability). 而软件在时刻 t 正常的概率称作**软件可用度**(software availability).

与硬件相比较，软件可靠性问题有其自身的特点. 软件不会因使用时间的增加而“老化”；软件的失效是由于人为设计或逻辑错误造成的；软件故障一旦发觉并排除后不会再出现同样的故障；软件不能通过冗余技术来提高其可靠性.

性质 33.11.2 软件开发的寿命周期：

(1) 目标阶段　为达到新开发系统所要求的功能，必须对相应的软件提出目标、要求及规范. 这是软件开发的出发点.

(2) 开发调试阶段　这一阶段中进行软件的编制、调试及改错.

(3) 论证阶段　这阶段是在开发者对软件质量表示满意之后进行的. 经过彻底的检验，目的在估计其可靠性.

(4) 运行阶段　软件通过论证后即交用户使用，并在使用过程中不断被检验.

(5) 维护阶段　在这阶段中及时改正运行中发现的错误，或随环境变化增加新的功能，或进行算法改进等.

(6) 软件更新阶段　随着系统环境的变化，软件不能完成各种新的功能时则需更新软件. 此时重新开始一个新的软件寿命周期.

模型 33.11.3 寿命周期的模型　在软件寿命周期中的不同阶段有不同的模型来描述. 如有描述软件错误的概率模型，用以预测或估计软件中剩余的错误数目或软件可靠性的增长；有在软件中人为加进错误，用以检验软件可靠性的模型；有与一定的费用相联系的软件发行时间的模型等. 目前以软件可靠性概率模型的研究为最多[100],[50].

杰林斯基-莫朗达(Jelinski-Moranda)在 1972 年提出了一个软件可靠性模型，是以后诸多模型的出发点. 模型的假定为：

(1) 一个软件系统原始错误的个数 N 是一未知常数.

(2) 软件相互故障之间的时间有独立的指数分布，其参数与

系统中残存的错误数目成正比，比例因子 λ 是常数.

(3) 当查出一个错误后，它被完全排除.

若记 $N(t)$ 为 $(0,t]$ 中发现的软件错误数，$N(o)=0$，$R(t)=N-N(t)$ 为到时刻 t 为止软件中残存的错误数，T_k 为发现第 k 个错误的时刻，$X_k=T_k-T_{k-1}$，$k=1,\cdots,N$，$T_0=0$. 则有

$$P\{R(t)=k\}=\binom{N}{k}e^{-\lambda kt}(1-e^{-\lambda t})^{N-k},k=0,\cdots,N; \quad (33.20)$$

$$ER(t)=Ne^{-\lambda t};$$

$$P\{X_k>t|T_{k-1}=s\}=P\{X_k>t\}=e^{-\lambda(N-k+1)t},k=1,\cdots,N.$$

若已观测到头上 n 次软件失效的时刻 $0=t_0<t_1<\cdots<t_n$，则未知参数 λ 及 N 可由对数似然函数

$$\log L=\sum_{k=1}^{n}(N-k+1)+n\lambda-\lambda\sum_{k=1}^{n}(N-k+1)(t_k-t_{k-1}) \quad (33.21)$$

求导获得. 即解 $\frac{\partial\log L}{\partial\lambda}=0$，$\frac{\partial\log L}{\partial N}=0$. 基于 λ,N 的估计即可对软件可靠性作出预测.

软件可靠性研究需要进一步深入的工作有：

(1) 不断完善模型. 即要建立更合适的模型来拟合已有的数据；引入或重新定义参数，以便于数据的收集、模型的数学表述以及统计估计及预测.

(2) 建立软件数据库.

34 随机模拟

34.1 引言

随机模拟(stochastic simulation)在历史上最早出现的例子，是有名的 Buffon 问题. 距今已有 200 多年了. 由于模拟试验工具的限制，一直未得到发展，随着生产管理及科学技术的发展，出现了许多复杂问题，如高维数学方程的求解、核物理中描述质点运动的微分方程、大系统的可靠性分析与模拟、地震波的模拟试验、医学、工程技术中的识别、诊断、系统工程的运筹优化、多元统计分析、排队网络等等. 这些问题用传统的物理试验或数学方法难以解决，而随机模拟的方法为这些问题的解决开辟了新的途径. 尤其是计算机的出现，为统计试验方法提供了强有力的模拟工具，才使这一方法得到了广泛的应用. 本章主要内容是：随机向量及随机过程的模拟；用统计试验方法求解数学、物理、工程技术与管理中的问题，如高维积分的计算、代数方程、微分方程求解等等.

34.2 均匀分布随机数的产生

34.2.1 均匀分布随机变量

设 X 的概率密度函数为

$$f(x)=\begin{cases}1, x\in[0,1],\\0, x\notin[0,1],\end{cases}$$

则 X 为[0,1]上均匀分布随机变量. 在计算机上，可产生随机变量

X 的抽样序列$\{x_n\}$，通常我们称 x_n 为[0,1]上均匀分布随机变量 X 的随机数，简称**均匀随机数**.

方法 34.2.1 分类 在计算机上，产生随机数的方法可分为三类：

(1) 把已有的随机数表输入机器(如美国 RAND 公司做的百万随机数表)；

(2) 用物理方法产生随机数；

(3) 利用数学方法产生随机数.

方法(3)发展快、简便易行，应用广泛，在此重点介绍.

34.2.2 伪随机数

定义 34.2.2 **伪随机数**(pseudorandom number)是指按照一定的计算方法产生的一列数，它们具有类似于均匀随机变量独立抽样值序列的性质. 由于这些数是依照确定性的算法产生的，存在着周期现象，但如计算方法选得恰当，它们便可近似于相互独立和近似于均匀分布，即它们能够通过数理统计中独立性检验和均匀分布检验. 因此人们称它们为伪随机数.

一般说来，伪随机数要借助于递推公式：

$$x_n = f(x_{n-1}, x_{n-2}, \cdots, x_{n-k}).$$

具体产生伪随机数的数学方法有迭代取中法，移位法及余数法等，下面仅介绍最常用的余数法. 令

公式 34.2.3 $y_{n+1} = \lambda y_n + c \pmod M$

其中初值 y_0，乘子 λ，增量 c 和模 M 取非负整数，当 $c=0$ 时

$$y_{n+1} = \lambda y_n \pmod M.$$

取

$$x_n = y_n M^{-1}, y_0 = a \ (a \text{ 为奇数}).$$

由构造方法可知：$0 \leqslant y_n \leqslant M$，$0 \leqslant x_n \leqslant 1$，故不同的 y_i(因之 x_i)至多只有 M 个，这表示序列$\{x_n\}$是有周期 L 的，$L \leqslant M$，每隔 L 个

不同的 x_i 后循环一次，于是 $\{x_n\}$ 不是真正的随机序列. 但若 L 充分大，在同一周期内的 $\{x_n\}$ 却可通过统计中独立性与均匀性检验，从而可视为服从[0,1]上均匀分布的伪随机数列. $\{x_n\}$ 完全依赖于参数 a、λ 及 M 的选择. 如何选择参数，目前主要依靠在计算机上作实验. 从文献中得知下列参数组较为适用：

(1) $a=1,\lambda=5^{17},M=2^{42}(L=2^{40}\approx10^{12})$；

(2) $a=1,\lambda=5^{13},M=2^{36}(L=2^{34}\approx2\times10^{10})$；

(3) $a=1,\lambda=7,M=10^{10}(L=5\times10^{7})$.

(参见文献[51],[103],[106],[109])

34.2.3 伪随机数的检验

为了确认所产生的随机数是否具有所要求的统计性质，通常要对它们进行一定的检验，下面介绍常用的统计检验.

方法 34.2.4 参数检验 为了检验随机数的分布参数抽样值与理论值的差异是否显著. 设产生的均匀随机数序列 $\{x_i\}$. 记一阶矩、二阶矩和方差的观察值为

$$\overline{X}=\frac{1}{N}\sum_{i=1}^{N}x_i,\quad \overline{X^2}=\frac{1}{N}\sum_{i=1}^{N}x_i^2,$$

$$s^2=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\frac{1}{2}\right)^2.$$

计算它们的期望值与方差分别为：

$$E(\overline{X})=\frac{1}{2},\quad D(\overline{X})=\frac{1}{12N},\quad E(\overline{X^2})=\frac{1}{3},$$

$$D(\overline{X^2})=\frac{4}{45N},\quad E(s^2)=\frac{1}{12},\quad D(s_N^2)=\frac{1}{180N}.$$

由中心极限定理知，当 N 充分大时

$$u_1=\frac{\overline{X}-E(\overline{X})}{\sqrt{D(\overline{X})}}=\sqrt{12N}\left(\overline{X}-\frac{1}{2}\right);$$

$$u_2 = \frac{1}{2}\sqrt{45N}\left(\overline{X^2} - \frac{1}{3}\right);$$

$$u_3 = \sqrt{180N}\left(s^2 - \frac{1}{12}\right).$$

渐近服从 $N(0,1)$ 分布. 据此可用统计量 u_1, u_2 及 u_3 分别对 $\overline{X}, \overline{X^2}$ 及 s^2 作假设检验.

方法 34.2.5 均匀性检验 方法如下:

把 $[0,1]$ 分为 k 个等区间, 按 $\{x_i\}$ 取值落在那个区间分为 k 组, 设有 n_j 个随机数在 $\left(\frac{j-1}{k}, \frac{j}{k}\right]$, $(j=1,2,\cdots,k)$. 零假设 H_0: X 落在每小区间概率 $P_j=\frac{1}{k}$. 当 H_0 成立时, 统计量

$$\chi^2 = \frac{k}{N}\sum_{j=1}^{k}\left(n_j - \frac{N}{k}\right)^2,$$

渐近服从 $\chi^2(k-1)$ 分布, 统计量

$$\sqrt{N}D_N = \frac{1}{\sqrt{N}}\max_{1<j<k}\left|\sum_{i=1}^{j}\left(n_i - \frac{N}{k}\right)\right|,$$

渐近服从科尔莫戈罗夫-斯米尔诺夫 Q_λ 分布. 取显著水平 $\alpha=0.05$, 当 $\sqrt{N}D_N>1.35$ 时, 拒绝 H_0.

方法 34.2.6 相关系数检验 相关系数为 0 是两个随机变量独立的必要条件. 这可用来作为前提检验随机数的独立性.

记前后相距为 j 的相关系数为

$$\bar{\rho}_j = \frac{1}{N-j}\left|\sum_{i=1}^{N-1}x_i x_{i+j} - (\bar{x})^2\right|/s^2, \qquad j=1,2,\cdots,$$

其中 $s^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{2}\right)^2$.

取 H_0: $\rho=0$, 对充分大的 $N(N-j>50)$ 统计量 $u=\bar{\rho}_j\sqrt{N-j}$ 渐近服从 $N(0,1)$ 分布.

方法 34.2.7 联列表检验 在 X-Y 平面上, 将单位正方形分

为 k^2 个相等的小正方形，把随机数序列$\{x_i\}$按出现的先后顺序两两分组. 如取(x_1,x_{l+1})，(x_2,x_{l+2})，…，(x_{N-l},x_N)，…，(x_N,x_l)作为点的 X-Y 坐标. 记落入小正方形(i,j)内的点数为 $n_{ij}(i,j=1,2,\cdots,k)$. 令 $\sum_{i=1}^{k} n_{ij}=n_{i\cdot}$，$n_{\cdot j}=\sum_{i=1}^{k} n_{ij}$，$N=\sum_i\sum_j n_{ij}$，取 H_0：$\{x_i\}$相互独立，当 H_0 成立时，统计量

$$\chi^2 = N\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{n_{ij}^2}{n_{i\cdot}\cdot n_{\cdot j}}-1\right)$$

渐近服从分布 $\chi^2[(k-1)^2]$.

34.3 随机变量及随机向量模拟

上节所造的随机数是[0,1]均匀分布随机数，简称为均匀随机数. 本节的目的是要在上节的基础上，构造出具有事先给定的分布函数 $F(x)$的随机数，这种数称为 **$F(x)$—随机数**，如 $F(x)$有密度函数 $f(x)$，那么也称它们为 **$f(x)$—随机数**. 设随机变量 η 的分布函数是 $F(x)$，所谓对 η 的模拟就是要构造若干个与 η 同分布的 $F(x)$——随机数；从概率论中易知，对于分布为 $F(x)$的随机数，都可从均匀随机数经过一些变换得到，可见均匀随机数的产生是随机模拟的基础. 下面，分别给出离散型随机变量、连续型随机变量及随机向量的模拟. 在本节记 X 为[0,1]上均匀分布的随机变量.

34.3.1 离散型随机数

设随机变量 η 的分布律为

$$P(\eta=a_n)=P_n,\quad n=0,1,2,\cdots,$$

其中 $0<P_n<1$，$\sum_{n=0}^{\infty}P_n=1$，$a_n\in R_1$. 令 $l_{-1}=0$，$l_k=\sum_{i=0}^{k}p_i$

($k \geqslant 0$). 显然成立事件间等式

$$(\eta = a_k) = (l_{k-1} < X \leqslant l_k), \tag{34.1}$$

$$k = 0,1,2,\cdots.$$

方法 34.3.1 模拟步骤如下：

(1) 取均匀随机数 $X = x_j$.

(2) 若 $\sum_{i=0}^{k-1} p_i < x_j \leqslant \sum_{i=0}^{k} p_i$,

则得 η 的第 j 个抽样值，$\eta_j = a_k$，$\{\eta_j\}$即为所求.

我们以泊松分布为例，说明对一些特殊的离散分布，如几何分布、二项分布、超几何分布等，从它们的概率意义出发，可设计和上述不同的一些模拟方法.

例 34.3.2 设随机变量 η 服从泊松分布，即

$P(\eta = n) = \dfrac{\lambda^n}{n!} e^{-\lambda}, n = 0,1,2,\cdots$ 且 $\lambda > 0$.

设均匀随机数序列$\{X_i\}$，令

$$(\eta = n) = \left(\prod_{i=0}^{n+1} X_i < e^{-\lambda} \leqslant \prod_{i=0}^{n} X_i\right)$$

或

$$\eta = \min\left\{n: \prod_{i=0}^{n+1} X_i < e^{-\lambda}\right\}.$$

方法 34.3.3 具体模拟步骤如下：

(1) 计算 $e^{-\lambda}$，令 $i = 0, S_i = 1$；

(2) 选取伪随机数 x_{i+1}；

(3) $S_{i+1} = S_i \cdot x_{i+1}$；

(4) 若 $S_{i+1} < e^{-\lambda}$转入(6)，否则转向(5)；

(5) 以 $i+1$ 代 i，S_{i+1}代 S_i 转向(2)；

(6) 令 $\eta = i$，则 i 即为所求的泊松分布随机数的一个取值，重复上述步骤即得所求.

34.3.2 逆变换方法

设已给分布函数为 $F(x)$，设 $F(x)$ 为连续函数，记它的反函数为 $F^{-1}(x)$，令 $\eta=F^{-1}(x)$，其中 x 为均匀随机变量，则易证，随机变量 η 的分布函数恰等于 $F(x)$. 可见 $F(x)$——随机数可自均匀随机数经过逆变换获得. 但实际上，仅当 F^{-1} 能明确地用解析式表达时才便于应用.

例 34.3.4 指数分布　指数分布函数 $F(x)=1-\mathrm{e}^{-\lambda x}, x\geqslant 0$. ($\lambda>0$ 为常数)

令 $X=1-\mathrm{e}^{-\lambda\eta}$. 解出 $\eta=-\frac{1}{\lambda}\ln(1-X)$，此时 $F^{-1}(X)=-\frac{1}{\lambda}\ln(1-X)$，因 X 是均匀分布，那么 $1-X$ 亦是均匀分布，因此上式可用 X 代 $1-X$，得

$$\eta=-\lambda^{-1}\ln X, \tag{34.2}$$

故只要产生均匀随机数 $\{x_i\}$，由变换 (34.2) 即得指数分布随机数 $\{\eta_i\}$.

例 34.3.5 韦布尔分布　其概率密度函数为

$$f(x)=\frac{\beta}{\alpha^\beta}x^{\beta-1}\mathrm{e}^{-(\frac{x}{\alpha})^\beta}, \text{且 } x\geqslant 0.$$

其中 $\alpha>0$ 及 $\beta>0$ 是分布的尺度和形状参数，位置参数设为 0. 其分布函数为 $F(x)=1-\mathrm{e}^{-(\frac{x}{\alpha})^\beta}, x\geqslant 0$.
令

$$F(\eta)=1-\mathrm{e}^{-(\frac{\eta}{\alpha})^\beta}=X$$

解得 $\eta=\alpha[-\ln(1-X)]^{1/\beta}$. 以 X 代 $1-X$ 得

$$\eta=\alpha[-\ln(1-X)]^{1/\beta} \tag{34.3}$$

利用 (34.3) 及均匀随机数 $\{x_i\}$ 可产生韦布尔分布的随机数 $\{\eta_i\}$.
对于逆变换法有时由于 $F(x)$ 的表达式复杂而难于求解. 下面介绍几种方法，不需要求解 $F(x)$ 的反函数.

34.3.3 筛选法

此法的实质是从许多均匀随机数选取一部分，使选出的那部分满足指定的 $f(x)$—随机数. 它可用于产生任意有界的随机数.

例 34.3.6 筛选法(sieve method)

设 η 为在 (a,b) 中取值的随机变量，其密度为 $f(x)$. (此时 $\int_a^b f(x)\mathrm{d}x=1$)，选取常数 α，使 $\alpha f(x)\leqslant 1, x\in(a,b)$，又设 X_1,X_2 为相互独立的 $[0,1]$ 上均匀随机变量，则 $Y_1=a+(b-a)X_1$ 是 $[a,b]$ 上均匀随机变量，规定筛选规则如下：如有 $X_2\leqslant\alpha f(Y_1)$ 则令 $\eta=Y_1$；否则舍去这一组 (X_1,X_2)，并重新产生一组 $[0,1]$ 上均匀随机变量(独立)，直到 $X_2\leqslant\alpha f(Y_1)$ 满足又产生一个 η 的样本. 如此重复，就得到 η 的样本列. 其统计密度视为 $f(x)$.

例 34.3.7 β分布的模拟　β分布随机变量的密度为

$$f(x)=\frac{\Gamma(\gamma+\lambda)}{\Gamma(\gamma)\Gamma(\lambda)}x^{\gamma-1}(1-x)^{\lambda-1}, 0\leqslant x\leqslant 1,$$

其中 $\gamma>0$ 及 $\lambda>0$ 为二参数，而

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x.$$

记 $M=\max f(x)=f(x)|_{x=(\gamma-1)/(\gamma+\lambda-2)}$.

模拟步骤如下：

(1) 由 γ 和 λ 计算 M；

(2) 产生[0,1]均匀分布伪随机数 x_1,x_2；

(3) 检验不等式，如

$$x_2M\leqslant f(x_1)=\frac{\Gamma(\lambda+\gamma)}{\Gamma(\lambda)\Gamma(\gamma)}x_1^{\gamma-1}(1-x_1)^{\lambda-1},$$

则取 $\eta=x_1$ 作为β分布的随机数；否则舍去 x_1,x_2，再转入步骤(2)，则 η 为β分布随机数.

34.3.4 分段逼近法

方法 34.3.8 分段逼近法(piecewise approximation method) 设分布函数集中在[a,b],将[a,b]分为 n 等分,分点为 $a=a_0<a_1<a_2<\cdots<a_n=b$. 令 $p_i=F(a_k)-F(a_{k-1})$. 将[0,1]也分为 n 个小区间,使第 i 个区间端点为 c_{i-1},c_i 且 $c_i-c_{i-1}=p_i$,现任取[0,1]上随机数 x,若 $c_{i-1}<X\leqslant c_i$,则令

$$\eta=a_{i-1}+(a_i-a_{i-1})\frac{x-c_{i-1}}{c_i-c_{i-1}},$$

当 n 充分大,可近似认为 η 是 $F(x)$——随机数.

若 $F(x)$是定义在无穷区间$(-\infty,+\infty)$上,那么对充分小的 $\varepsilon>0$,总可选取[a,b]使 $F(b)-F(a)\geqslant 1-\varepsilon$. 然后在[$a,b$]上运用上述方法.

34.3.5 正态分布 $N(0,1)$随机数的产生

正态分布使用频繁,这里介绍三种产生正态分布随机数的方法.

方法 34.3.9 筛选法 产生步骤如下:

(1) 产生[0,1]上均匀随机变量 X_1,X_2.

(2) 计算 $V_1=2X_1-1,V_2=2X_2-1$ 及 $s=V_1^2+V_2^2$;

(3) 若 $s\leqslant 1$,转向步骤(4),否则转向(1);

(4) 计算 $A=\left(-2\frac{\ln s}{s}\right)^{\frac{1}{2}},\xi=V_1A,\eta=V_2A$;

则ξ,η 为 $N(0,1)$的随机变量. 此法精度高,由 Box-Mulles 提出.

方法 34.3.10 变换法(transformation method) 设 X_1,X_2 是[0,1]上均匀随机变量,相互独立,则

$$Z_1=\sqrt{-2\ln X_1}\cos 2\pi X_2,$$

$$Z_2 = \sqrt{-2\ln X_1}\sin 2\pi X_2.$$

易证 Z_1,Z_2 是相互独立的 $N(0,1)$ 分布的随机数，此法计算量一般较大.

方法 34.3.11 近似法(approximation method) 设 $X_1,X_2,\cdots,X_n$ 是[0,1]上均匀分布，相互独立的随机变量，由中心极限定理知，当 n 充分大时，

$$Z = \sum_{i=1}^{n}\left(X_i - \frac{1}{2}\right)\Big/ n \cdot \frac{1}{12}$$

近似于 $N(0,1)$ 分布. 通常取 $n=12$. 即 $Z=X_1+X_2+\cdots+X_{12}-6$ 作为 $N(0,1)$ 的近似.

34.3.6 随机向量的模拟

设 n 维随机向量 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$，它的概率密度 $f(x_1,x_2,\cdots,x_n)$ 可写成

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)f(x_2|x_1)f(x_3|x_1,x_2)\cdots f(x_n|x_1,x_2,\cdots x_{n-1}), \quad (34.4)$$

其中 $f_1(x_1)$ 是 ξ_1 的边缘分布，$f(x_k|x_1,\cdots,x_{k-1})$ 是在 $\xi_1=x_1\cdots\xi_{k-1}=x_{k-1}$ 条件下，ξ_k 的条件分布密度. (34.4)启示模拟 ξ 的方法，即先选出 $f_1(x_1)$——随机数 y_1，以 y_1 代入 $f(x_2|x_1)$ 中的 x_1，得一给定 y_1 的一维密度 $f(x_2|y_1)$，再造 $f(x_2|y_1)$——随机数 $y_2,\cdots$，如 $y_1,y_2,\cdots y_{k-1}$ 已确定，再造出 $f(x_k|y_1,\cdots y_{k-1})$——随机数 y_k，所得 $(y_1,\cdots,y_n)$ 是 $(\xi_1,\xi_2,\cdots\xi_n)$ 的模拟.

对离散型随机向量的模型可类似作出.

方法 34.3.12 n 维正态模拟的简便方法 对 n 维正态分布的随机向量有下列的简便方法. 设 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$ 联合密度

$$f(x_1,x_2,\cdots,x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{B}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{X}^{-1}},$$

其中 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$，$\boldsymbol{B}=(b_{ij})$ 为 n 维正定对称矩阵，易知 $E\xi_i$

$=0, E\xi_i\xi_j=b_{ij}$. 现取 n 个相互独立的一维 $N(0,1)$ 正态分布随机变量 $X_1, X_2, X_3, \cdots, X_n$，作变换

$$\begin{cases} Y_1 = a_{11}X_1, \\ Y_2 = a_{21}X_1 + a_{22}X_2, \\ \vdots \\ Y_n = a_{n1}X_1 + a_{n2}X_2 + \cdots + a_{nn}X_n, \end{cases} \tag{34.5}$$

其中 a_{ij} 满足

$$b_{ij} = \sum_{k=1}^{s} a_{ik}a_{jk} \quad (s = \min(i,j)), \tag{34.6}$$

即有

$$a_{11} = b_{11}^{\frac{1}{2}}, a_{21} = \frac{b_{12}}{b_{11}^{\frac{1}{2}}}, a_{22} = \left(\frac{b_{11}b_{22} - b_{12}^2}{b_{11}}\right)^{\frac{1}{2}}. \tag{34.7}$$

……

一般，如已求出 $a_{11}; \cdots; a_{k1}, a_{k2}, \cdots, a_{kk}$，可由(34.6)求出 $a_{k+1,1}$，$a_{k+1,2}, \cdots, a_{k+1,k+1}$. 故 (a_{ij}) 由 $\boldsymbol{B}=(b_{ij})$ 决定. 这样从 n 维 $N(0,I)$ 的伪随机数矢量 $(x_1, \cdots, x_n)$ 得到模拟 ξ 的一个伪随机数序列 $(y_1, \cdots, y_n)$.

34.4 随机过程的模拟

以下介绍常用的几种随机过程的模拟.

34.4.1 马尔可夫链的模拟

方法 34.4.1 离散时间马尔可夫链的模拟 设 $\{Y_n, n \geqslant 0\}$ 是马尔可夫链，它的一步转移矩阵 $\boldsymbol{P}=(p_{ij})$，状态空间 $S=\{0,1,2,\cdots\}$. 初始分布 $\boldsymbol{\pi}=(\pi_0, \pi_1, \cdots, \pi_i, \cdots)$. 模拟步骤如下：

(1) 任取[0,1]均匀随机数 x_0，如 $\sum_{i=0}^{k_0-1}\pi_i < x_0 \leqslant \sum_{i=0}^{k_0}\pi_i$，取

$y_0=k_0$；

（2）再取[0,1]均匀随机数 x_1，如 $\sum\limits_{i=0}^{k_1-1} p_{k_0 i}<x_1 \leqslant \sum\limits_{i=0}^{k_1} p_{k_0 i}$，取 $y_1=k_1$；

（3）一般地，设 $y_n=k_n$，已确定，再取[0,1]上均匀随机数 x_{n+1}，如 $\sum\limits_{i=0}^{k_{n+1}-1} p_{k_n i}<x_{n+1} \leqslant \sum\limits_{i=0}^{k_{n+1}} p_{k_n i}$，取 $y_{n+1}=k_{n+1}, \cdots$.

这样模拟得到的 $\{Y_n, n \geqslant 0\}$ 即所求的马尔可夫链的一条轨道.

方法 34.4.2 连续时间马尔可夫链模拟 设连续时间马尔可夫链 $\{Y_t, t \geqslant 0\}$ 其转移率矩阵 $\boldsymbol{Q}=(q_{ij})$，初始分布

$$\boldsymbol{\pi}=(\pi_0, \pi_1, \cdots, \pi_i, \cdots)，\text{状态空间 } S=\{0,1,2, \cdots\},$$

记 $q_i=-q_{ii}, p_{ij}=\dfrac{q_{ij}}{q_i}, j \neq i, p_{ii}=0$.

模拟步骤 设 $\{U_n, n \geqslant 0\}\{V_n, n \geqslant 1\}$ 为相互独立，[0,1]上均匀随机数序列.

（1）如 $\sum\limits_{i=0}^{k_0-1} \pi_i<U_0 \leqslant \sum\limits_{i=0}^{k_0} \pi_i$，取 $Z_0=k_0$，且取 $\tau_1=-q_{k_0}^{-1} \ln V_1$.

（2）如 $\sum\limits_{i=0}^{k_1-1} p_{k_0 i}<U_1 \leqslant \sum\limits_{i=0}^{k_1} p_{i_{k_0} i}$，取 $Z_1=k_1$，且取 $\tau_2=-q_{k_1}^{-1} \ln V_2$.

（3）如 $\sum\limits_{i=0}^{k_{n+1}-1} p_{k_n i}<U_{n+1} \leqslant \sum\limits_{i=0}^{k_{n+1}} p_{k_n i}$，取 $Z_{n+1}=k_{n+1}$，且取 $\tau_{n+1}=-q_{k_n}^{-1} \ln V_{n+1}$.

（4）令 $\tau_0=0$，当 $t \in\left[\sum\limits_{i=0}^{n} \tau_i, \sum\limits_{i=0}^{n+1} \tau_i\right)$，定义 $Y_t=k_n$，

则 $\{Y_t, t \geqslant 0\}$ 即为所求的马尔可夫链.

注 34.4.3 若 $p_{i i+1}=1, p_{ij}=0, j \neq i+1,(i=0,1,2, \cdots)$，则 $\{Y_t, t \geqslant 0\}$ 是泊松过程.

34.4.2 宽平稳序列的模拟

方法 34.4.4 宽平稳序列的模拟 设宽平稳序列$\{X(n),n\geqslant\varphi\}$,$EX(n)=0$,协方差函数为$R_n$,令

$$\begin{cases}X(1)=c_1\eta_1+c_2\eta_2+\cdots+c_n\eta_n,\\X(2)=c_1\eta_2+c_2\eta_3+\cdots+c_n\eta_{n+1},\\\vdots\\X(k)=c_1\eta_k+c_2\eta_{k+1}+\cdots+c_n\eta_{n+k-1},\\\vdots\\X(n)=c_1\eta_n+c_2\eta_{n+1}+\cdots+c_n\eta_{2n-1},\end{cases}\tag{34.8}$$

其中$\{\eta_i\}$为任意满足$E\eta_i=0$,方差$D\eta_i=R_0$及$E\eta_i\eta_j=0(j\neq i)$的随机变量,通常可取$\{\eta_i\}$为独立的$[-h,h]$上均匀随机数,方差为$\frac{h^2}{3}=R_0$.或取为独立的$N(0,R_0)$随机数,$c_i(1\leqslant i\leqslant u)$可由下列方程组

$$R_{k-1}=R_0(c_1c_k+c_2c_{k+1}+\cdots+c_{n-k+1}c_n),k=1,\cdots,n.\tag{34.9}$$

解出代入(34.8)即得所需的平稳序列$\{Z(n),n\geqslant 0\}$.

对于宽平稳过程模拟,可用宽平稳序列逼近.

34.4.3 平稳正态过程的模拟

方法 34.4.5 平稳正态马尔科夫过程的模拟 设$\{X(t),t\geqslant 0\}$是平稳正态马尔可夫过程,其协方差函数为$B_1(\tau)=\sigma^2e^{-\beta|\tau|}$($\sigma>0,\beta>0$常数)且$EX(t)=0$.在电子计算机上模拟$\{X(t)\quad t\geqslant 0\}$,只能得到有限或可列多个随机变量,记$X_n=X(n\Delta t),n=1,2,\cdots$($\Delta t$为常数).应用上节结果,模拟$\{X_n\quad n\geqslant 0\}$步骤如下:

(1) 产生相互独立的$N(0,1)$分布的随机数$\{y_n\quad n\geqslant 1\}$.

(2) 计算

$$X_1=\sigma y_1, X_k=e^{-\beta\Delta t}X_{k-1}+\sigma\sqrt{1-e^{-2\beta\Delta t}}y_k, 1\leqslant k,$$

$\{X_n, n\geqslant 1\}$即为所求.

方法 34.4.6 平稳正态过程的模拟 设$\{X(t), t\geqslant 0\}$为平稳正态过程(22.4.1)其协方差函数为 $B_2(\tau)=\sigma^2 e^{-\beta|\tau|}\cos\omega\tau(\sigma>0, \beta>0, \omega>0$ 常数$)EX(t)=0$.

模拟步骤 由方法 34.4.5 中模拟方法产生两个独立的平稳正态马尔可夫过程 $X_1(t)$及 $X_2(t)$,使其 $EX_1(t)=EX_2(t)=0$,协方差函数均等于 $\sigma^2 e^{-\beta|\tau|}$;再计算 $X(t)=X_1(t)\cos\omega t+X_2(t)\sin\omega t$,则$\{X(t), t\geqslant 0\}$即为所求的过程.

34.4.4 维纳过程的模拟

方法 34.4.7 维纳过程的模拟 设$\{X(t), t\geqslant 0\}$为维纳过程,取离散化步长为 Δt,记 $X_n=X(n\Delta t)$. 模拟$\{X_n, n\geqslant 0\}$的步骤如下:

(1) 产生相互独立的 $N(0,1)$随机数$\{W_n, n\geqslant 1\}$.

(2) 计算 $X_1=\sigma\sqrt{\Delta t}W_1, X_n=X_{n-1}+\sigma\sqrt{\Delta t}W_n(X_0=0)$,则$\{X_n, n\geqslant 0\}$即是所求.

34.5 统计试验法

统计试验方法又称蒙特卡罗(Monte Carlo)方法,是一类通过随机变量的统计试验、随机模拟,求解数学物理、工程技术及经济管理近似解的数值方法.

为说明统计试验法的基本思想,这里通过一个简单的积分计算问题作为例子.

34.5.1 积分的计算

例 34.5.1 计算积分 定积分计算,设 $0<g(x)<1$,计算

积分$\int_0^1 g(x)\mathrm{d}x \triangleq J$.

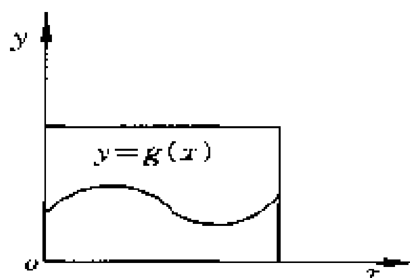


图 34.1

模拟的基本思想　积分J等于由曲线$y=g(x)$、x轴、y轴及$x=1$为边界围成的面积(见图 34.1),为用蒙特卡罗(Monte Carlo)方法计算积分J,我们在单位正方形($0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1$)中,随机地投掷一点,即它的两坐标$(x,y)$都是$[0,1]$均匀分布的随机变量且相互独立. 易知,这一随机点$(x,y)$落在所求区域中(即$y\leqslant g(x)$)的概率$P$正好等于$J$,即

$$J=\int_0^1 g(x)\mathrm{d}x=P$$

若重复上述试验n次,即随机地投掷n个点. 令

$$\eta_i=\begin{cases}1, \text{如第 } i \text{ 次投掷点}(x,y)\text{满足 } y\leqslant g(x), \\ 0, \text{如第 } i \text{ 次投掷点}(x,y)\text{满足 } y>g(x),\end{cases}$$

则$\sum_{i=1}^{n}\eta_i$表示投掷点落在所求区域的次数,$Q_n=\frac{1}{n}\sum_{i=1}^{n}\eta_i$表示相应的频率. 由强大数定律有$P(\lim_{n\to+\infty}Q_n=P)=1$,故当$n$充分大时,以概率 1,有

$$J=P\approx Q_n=\frac{1}{n}\sum_{i=1}^{n}\eta_i,$$

且$EQ_n=J$. 因而我们用Q_n作为所求量J的估计值,Q_n可用试验求得. 由上说明,人为构造一种概率模型,使它的统计试验的某些特征数(如频率、样本均值等)作为所求量的估计值,这便是蒙特卡

罗方法的基本思想.

精度　以 Q_n 近似 J，如何考虑它的精确度 ε，通常要看 Q_n 的方差，$DQ_n=\frac{1}{n}J(1-J)$. 或用样本方差 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(\eta_i-Q_n)^2$ 作估计，也可用概率来刻画：给定 n 与 $\varepsilon>0$，求 $P(|Q_n-J|<\varepsilon)=?$ 反之，给定 ε 及 $0<\alpha<1$. 问应取多大的 n，使 $|Q_n-J|<\varepsilon$ 的概率不小于 $(1-\alpha)$，即要求 n 使

$$P(|Q_n-J|<\varepsilon)\geqslant 1-\alpha. \qquad (34.10)$$

由中心极限定理及上式可得

$$P(|Q_n-J|<\varepsilon)\approx\Phi\left[\varepsilon\sqrt{\frac{n}{J(1-J)}}\right]-\Phi\left[-\varepsilon\sqrt{\frac{n}{J(1-J)}}\right]\geqslant 1-\alpha,$$

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u$. 令 $\Phi(x_{\frac{\alpha}{2}})-\Phi(-x_{\frac{\alpha}{2}})=1-\alpha$，于是给定模拟精度 ε，所需的模拟次数由下式给出.

$$n\geqslant(x_{\frac{\alpha}{2}})^2\frac{J(1-J)}{\varepsilon^2}. \qquad (34.11)$$

(34.11)中的 J，可用试算若干次后的估计值代替.

步骤　用统计试验法求解一实际问题的大致步骤如下：

(1) 根据问题，构筑模拟概型，使它的若干数字特征恰好是所要求的量；

(2) 根据问题的特点，设计及应用降低方差的各种方法，加速模拟的收敛速度；

(3) 给出概型中各种不同分布随机变量或随机过程的抽样方法；

(4) 统计处理模拟结果，给出问题的解及解的精度估计.

方法 34.5.2　平均值方法　在模拟中为何提高精度而不增加模拟次数？或降低 n 又能保证精度？下面用另一平均值方法计算积分 $Q=\int_a^b g(x)\mathrm{d}x$ 为例说明这个问题. 取一列相互独立，同在

$[a,b]$ 上均匀分布的随机变量 $\{\xi_i\}$，则有

$$Eg(\xi_i) = \frac{1}{b-a}\int_a^b g(x)\mathrm{d}x = \frac{Q}{b-a}, \qquad (34.12)$$

取 $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} g(\xi_i)\cdot(b-a)$ 作为 Q 的估计，则有 $E\hat{Q}_n = Q$，及

$$P(\lim\hat{Q}_n = Q) = 1.$$

方差为

$$DQ'_n = \frac{(b-a)^2}{n}Dg(\xi_i)$$

$$= \frac{1}{n}\left[(b-a)\int_a^b g^2(x)\mathrm{d}x - \left(\int_a^b g(x)\mathrm{d}x\right)^2\right]. \qquad (34.13)$$

注 34.5.3 不同模拟方法的比较 若要比较用不同统计试验方法计算 $\int_a^b g(x)\mathrm{d}x$ 时的方差. 则由(34.13)知，取 $k>1$，如以 $\frac{g(x)}{k}$ 代替(34.13)式中的 $g(x)$，方差就降低 k^2 倍. 方差小则精确度高，并使运算次数减少，计算速度加快，故在蒙特卡罗方法中，如何降低方差是一重要问题. 易证，当 $a=0, b=1, g(x)\leqslant 1$ 时，$D\hat{Q}_n \leqslant DQ_n$，即用方法 34.5.2 较有效.

方法 34.5.4 多重积分计算 上述二种方法易于推广应用于计算多维重积分. 记 $x=(x_1,\cdots,x_m)$ 为 m 维空间 R^m 中 m 点，$g(x)$ 为 m 维函数，它在闭域 $D\subset R^m$ 上可积，为计算

$$J = \int_D g(x)\mathrm{d}x = \int_D\cdots\int g(x_1,\cdots,x_m)\mathrm{d}x_1\cdots\mathrm{d}x_m,$$

取 $[a,b]$，$b-a$ 充分大，使 $D\subset[a,b]\times[a,b]\times\cdots\times[a,b]\triangleq[a,b]^m$ 取 m 个独立，在 $[a,b]$ 上均匀分布的随机变量 $\xi_1,\xi_2,\cdots,\xi_m$，记 $\eta=(\xi_1,\xi_2,\cdots,\xi_m)$.

多重积分计算的模拟步骤：

(1) 产生一列独立的，在 $[a,b]^m$ 上均匀分布的 m 维随机向量 $\{\eta^{(i)}\}$，$\eta^{(i)}=(\xi_1^{(i)},\xi_2^{(i)},\cdots,\xi_m^{(i)})$.

(2) 如 $\eta^{(i)}\in D$,就取 $\eta^{(i)}$,并记 $\hat{\eta}^{(i)}=\eta^{(i)}$;如 $\eta^{(i)}\bar{\in} D$,就舍弃 $\eta^{(i)}$.

如此共选取 n 个 $\hat{\eta}^{(i)}$,记 $J_n=|D|\cdot\frac{1}{n}\sum_{i=1}^{n}g(\hat{\eta}^{(i)})$,其中$|D|$为 D 的 m 维体积,取 J_n 作为 J 的估计值.

易证 $EJ_n=J$,且 $P(\lim_{n\to\infty}J_n=J)=1$.

34.5.2 线性方程组的解法

方法 34.5.5 一般线性方程组的解法 为解线性方程组

$$\sum_{j=1}^{n}c_{ij}x_j=d_i,\quad i=1,2,\cdots,n. \tag{34.14}$$

考虑二次型

$$\begin{aligned}Q(\boldsymbol{x})&=Q(x_1,x_2,\cdots,x_n)\\&=\sum_{i=1}^{n}a_i\Big(\sum_{j=1}^{n}c_{ij}x_j-d_i\Big)^2,\end{aligned} \tag{34.15}$$

这里 $a_i>0$,常数,$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$. 显然,求(34.14)的解等价于求 $Q(\boldsymbol{x})$的最小点$(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})$. 设任给常数 $A>0$,记 $E=\{x=(x_1,\cdots,x_n):Q(x_1,\cdots,x_n)\leqslant A\}$是 n 维椭球. 作 n 维立方体$[a,b]^n$使 $E\subset[a,b]^n$. 设 $\xi_1,\xi_2,\cdots,\xi_n$ 是$[a,b]$上均匀分布 n 个独立的随机变量,记 $\boldsymbol{Z}=(\xi_1,\xi_2,\cdots,\xi_n)$为 n 维随机向量. 求(34.14)的解$(x_1^{(0)},\cdots,x_n^{(0)})$步骤如下:

(1) 取 l 个独立的 n 维随机向量 $\boldsymbol{Z}^{(1)},\boldsymbol{Z}^{(2)},\cdots,\boldsymbol{Z}^{(l)}$.

(2) 若 $\boldsymbol{Z}^{(1)}\in E$,取 $\boldsymbol{Z}^{(1)}$,记为 $\boldsymbol{X}^{(1)}$;如 $\boldsymbol{Z}^{(1)}\bar{\in} E$,就舍弃 $Z^{(1)}$,如此重复共选取 m 个$(m\leqslant l)$,不妨记为 $\hat{\boldsymbol{X}}^{(1)},\hat{\boldsymbol{X}}^{(2)},\cdots,\hat{\boldsymbol{X}}^{(m)}$,其中 $\hat{\boldsymbol{X}}^{(i)}=(x_1^{(i)},x_2^{(i)},\cdots,x_n^{(i)})$,$i=1,2\cdots,m$.

(3) 取 $\hat{X}_j(m)=\frac{1}{m}\sum_{i=1}^{m}x_j^{(i)}$,

作为 $x_j^{(0)}$ 的估计(即以 $\hat{x}_j(m)$作为 $x_j^{(0)}$ 的近似),则有 $E\hat{X}_j(m)=$

$x_j^{(0)}, j=1,2,\cdots,n$，且 $\quad P(\lim_{m\to\infty}\hat{X}_j(m)=x_j^{(0)})=1.$

方法 34.5.6 某些特殊的线性方程组的解法 设方程组为

$$X = HX + A, \tag{34.16}$$

其中 $X=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$H=(h_{ij})_{n\times n}$. $A=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$，设 H 矩阵，A 向量为已知，求解未知向量 X. 记 $\|H\| = \max_i \sum_{j=1}^{n} |h_{ij}|$.

下面在假定 $\|H\|<1$ 条件下用蒙特卡罗方法求解(34.16)，我们定义一个马尔可夫链：任取一矩阵 $P=(p_{ij})_{n\times n}$ 满足 $p_{ij}\geqslant 0$，$\sum_{j=1}^{n} p_{ij} < 1$，$p_{ij}>0$ 若 $h_{ij}\neq 0$，令 $P_i = 1-\sum_{j=1}^{n} p_{ij}$(如 $a_i=0$ 则取 $p_i>0$). 模拟一马尔可夫链，使它的一步转移矩阵为 $P=(p_{ij})$.

模拟步骤：

(1) 仿照方法 34.4.1，模拟以 $P=(p_{ij})$ 为转移概率矩阵的马尔可夫链.

(2) 设模拟的运动轨道 T 记为

$$T: i\to i_1\to i_2\to\cdots\to i_k$$

表示从 i 出发，经 k 步后停在状态 i_k，($k\geqslant 0, i=i_0$). 定义一个轨道函数

$$V(\tau) = \begin{cases} \dfrac{h_{ii_1}}{p_{ii_1}} \quad \dfrac{h_{i_1i_2}}{p_{i_1i_2}} \quad \cdots \quad \dfrac{h_{i_{k-1}i_k}}{p_{i_{k-1}i_k}} \quad \dfrac{a_{i_k}}{p_{i_k}}, & \text{如 } k>0, \\ a_i/p_i, & \text{如 } k=0. \end{cases} \tag{34.17}$$

(3) 独立地按上述方法模拟 m 次(从 i 状态出发)，得到 n 个轨道 T_l，按(34.17)计算相应的轨道函数 $V(T_l)$，$l=1,2,\cdots,m$.

(4) 以 $\bar{x}_i(m) = \dfrac{1}{m}\sum_{l=1}^{m} V(T_l)$ 作为方程组(34.16)解 x_i^0 的估计值.

易证 $X_i^{(0)} = E_i(\bar{x}_i(m))$，$E_i$ 表示自 i 出发的条件数学期望.

此方法的特点是每个 $x_i^{(0)}$ 可单独求之，无需同时算其他的

$x_j^{(0)}$. 本方法亦可求解某些类型的积分方程.

34.5.3 某些偏微分方程的解法

这里仅以用蒙特卡罗方法求解拉普拉斯方程作为例子，说明它们的基本思想与步骤.

方法 34.5.7 求解平面上的拉普拉斯方程 设方程

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,(x,y)\in G, \tag{34.18}$$

$$u(x,y)=f(x,y),(x,y)\in \Gamma=\partial G.$$

要求 $u(x,y)=u(a)$ 当 $a\in G$ 时，使 $u(a)$ 满足方程(34.18). 当 $a\in G$，而在 G 的边界 Γ 上满足 $u(a)=f(a),a\in F$，这里 f 为已知函数.

用电子计算机求微分方程的数值解，一般要用网络法离散化微分方程变为对应的差分方程，设 G 上的网络步长为 h，G 离散化后的边界为 Γ_h，G 中的结点记为 G_h，对应方程(34.18)的差分方程为

$$\begin{cases}u(a)=\dfrac{1}{4}\sum\limits_{i=1}^{4}u(a_i), & a\in G_h,\\ u(a)=f(a), & a\in \Gamma_h.\end{cases} \tag{34.19}$$

其中 $a_i(i=1,2,3,4)$为结点 a 的四个相邻结点(如图 34.2).

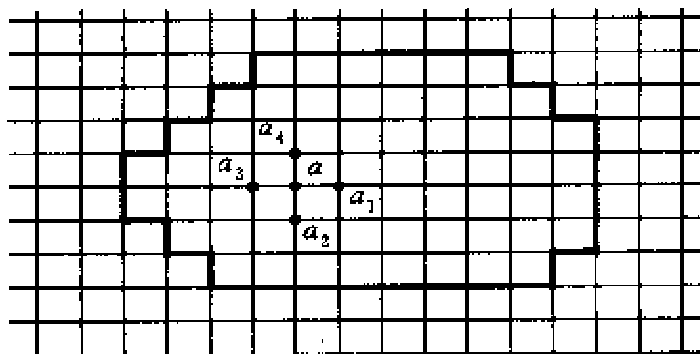


图 34.2

方法 34.5.8 模拟概型如下：

(1) 考虑自结点 $a\in G_h$ 出发在平面网络结点上随机游动的质点 M，它下一步到达四相邻结点的概率各为 1/4，再下一步又同样以 1/4 的概率到达四个相邻结点之一，如此继续，直至首次到达边界 Γ_h 时便被吸收而停止运动. 记下首次到达边界的点 $b\in\Gamma_h$.

(2) 独立地重复上述随机游动 n 次，每次从点 a 出发，第 i 次首达边界的点记为 $b_i\in\Gamma_a(i=1,\cdots,n)$.

(3) 以 $\hat{u}_n(a)=\frac{1}{n}\sum_{i=1}^{n}f(b_i)$ 作为 $u(a)$ 的估计值，易证 $E(\hat{u}_n(a))=u(a)$，且当 $a\in\Gamma_h$ 时，$\hat{u}_n(a)=f(a)$. 由大数定律有 $P(\lim_{n\to\infty}\hat{u}_n(a)=u(a))=1$. 这说明当 n 充分大时，$\hat{a}(n)$ 是满足方程 (34.19) 的近似解. 注意，(34.19) 的解实际上依赖于 h，但当 h 充分小时，它近似于 (34.18) 的解.

(参见[51]，[2]，[109]).

35 马尔可夫决策过程

35.1 引言

马尔可夫决策过程(Markov decision processes,简称 MDP)是研究由一系列决策(选取行动、措施、方案或参数等)控制的随机动态系统的优化问题.至今已形成为介于运筹学与应用概率统计之间的新分支.

MDP 已成功地应用于许多领域.例如设备的最优维修更换问题,库存、排队系统、可靠性工程、控制、搜索、管理调度、序贯抽样及质量控制等的优化问题,以及自然资源合理开采、利用、经济领域中的投资、信贷及消费的最优控制,城市管理、森林管理、渔业管理等等.

MDP 要求系统可周期地(离散时间)被观察(或连续地被观察),在每个观察时刻,决策者根据系统状态,从可用行动集中选用一个(即作决策)并施加于系统,则有两事件发生:

(1) 系统状态转移律的确定,并具有马尔可夫(无后效)性.

(2) 可得到一定的报酬(或费用),通常它仅与该时刻的状态及采用的行动有关.

决策者在各个时刻选取行动的目的是使系统运行的全过程,在某种给定准则下,达到最优运行.

下面先叙述离散时间 MDP 模型的基本构成.

35.2 离散时间MDP的基本要素

模型 35.2.1 一个离散时间MDP模型，由如下5要素组成 $\langle S,A,q,r,V\rangle$.

S 为**状态空间**(state space)，它是系统所有可能状态的非空集，为简单计，设 $S=\{1,2,\cdots,m\}$ 为有限集.

A 为**行动空间**(action space)，$A(i)$ 为状态 i 可用的行动集 $(i\in S)$，$A(i)\subset A$. A 为有限集.

q：$Q=(q_{ij})$ 为**转移概率矩阵**(transition probability matrix) $q_{ij}(a)$ 为在任一时刻 $t(t=0,1,2,\cdots)$ 系统在状态 i 采取行动 a 的条件下，于 $t+1$ 时刻转移到状态 j 的概率，它与系统在 t 前的历史无关(即马尔可夫转移律)也与时刻 t 无关(时齐性). 具有

$$q_{ij}(a)\geqslant 0,\ \sum_{j\in s}q_{ij}(a)=1,\ i,j\in s,\ a\in A.$$

r 为**报酬函数**(reward function)，令 $\Gamma=\{(i,a):a\in A,i\in s\}$，$r:\Gamma\to R$. R 为实数集.

$r(i,a)$ 表示在任意 t 时，系统于状态 i 选用行动 a 在一个周期所获得的报酬，称为报酬函数.

V 为**准则**(criteria)，或称**目标函数**(objective function) 记 Π 为全体**策略集**(严格定义见节35.3)，则 $V:\Pi\times s\to R$，目标函数.

给定一策略 $\pi\in\Pi$，状态 $i\in S$，$V(\pi,i)$ 表示 $t=0$ 时从状态 i 出发的条件下用策略 π 的目标值，它是选择和衡量策略 π 优劣的标准. 在节35.4将具体给出几个常用的目标函数. 当 $\langle S,A,q,r,V\rangle$ 确定时，就给定了MDP的一个具体模型.

35.3 策略

定义 35.3.1 设 $f:S\to A$，则称 f 为**决策函数**(action

function)或称**确定性决策规则**.

对 $i\in S$，$f(i)$表示在状态 i 选用行动，$f(i)\in A$. 全体决策函数之集记作 F，即 $F=\{f:S\to A\}$.

定义 35.3.2 决策函数序列 $\pi=\{f_0,f_1,\cdots,f_t,\cdots\}$称 π 为**确定性马尔可夫策略**(deterministic Markov policy)，其中 $f_t\in F$ 表示 t 时**决策函数**($t\geqslant 0$，下同).

全体确定性马尔可夫策略集记作 Π_M^d.

定义 35.3.3 若 $\pi=\{f,f,\cdots,f,\cdots\}$，其中 $f\in F$，则称 π 为**确定性平稳策略**(deterministic stationaary policy)，记作 $f^\infty(f\in F)$. 全体确定性平稳策略集记作 Π_s^d.

为引进一般随机策略，先定义"历史".

定义 35.3.4 记 $h_t=\{i_0,a_0,i_1,a_1,\cdots,i_t,a_t\}$，称 h_t 为系统直到 t 时刻的一个**历史**(history)，其中 $i_n\in S$，$a_n\in A(0\leqslant n\leqslant t)$.

记 $H_t=\{h_t\}$为系统直到时刻 t 的历史集.

注 35.3.5 记 $\pi_t(\cdot\mid h_{t-1},i_t)$为 A 上的概率分布. 即 $\pi_t(a\mid h_{t-1},i_t)$为已知历史 h_{t-1}及 t 时状态 i_t 条件下选用 a 的概率.

显然，$\pi_t(a\mid h_{t-1},i_t)\geqslant 0$，且 $\sum_{a\in A}\pi_t(a\mid h_{t-1},i_t)\leqslant 1$，对任意 $a\in A$，$h_{t-1}\in H_{t-1}$，$i_t\in S$，$t\geqslant 0$.

定义 35.3.6 $\pi=\{\pi_0,\pi_1,\pi_2,\cdots,\pi_t,\cdots\}$，称 π 为**策略**，即 π 是一串选用行动规则的条件概率分布序列族(π_t 对每一 $h_{t-1}\in H_{t-1}$ 及 $i_t\in S$ 均有定义).

定义 35.3.7 $\pi=\{\pi_0,\pi_1,\cdots,\pi_t,\cdots\}$，且 π_t 满足

$\pi_t(a\mid h_{t-1},i_t)=\pi_t(a\mid i_t)$，对所有 $a\in A$，

$$h_{t-1}\in H_{t-1},\quad i_t\in S,\quad t\geqslant 0,$$

即 π_t 与历史无关，则称 π 为**马尔可夫策略**. 全体马尔可夫策略之集记作 Π_M，称为**马尔可夫策略类**.

若 $\pi_t(\cdot\mid i_t)$为退化分布. 即存在 f_t 使 $\pi_t(f_t(i_t)\mid i_t)=1(t\geqslant 0)$，

则记 $\pi=\{f_0,f_1,\cdots,f_t,\cdots\}\in\Pi_M^d$.

定义 35.3.8 设 $\pi=\{\pi_0,\pi_1,\cdots,\pi_t,\cdots\}\in\Pi_M$,且 $\pi_t=\pi_0$,对 $t\geqslant 0$,称 π 为**平稳策略**,全体平稳策略之集记作 Π_s,称为**平稳策略类**.

命题 35.3.9 $\Pi_s^d\subset\Pi_M^d\subset\Pi_m\subset\Pi$,$\Pi_s^d\subset\Pi_s\subset\Pi_M\subset\Pi$.

不难证明 F 包含的元素数目与 Π_s^d 包含的元素数目一样多.

35.4 目标函数

记 Y_t 及 Δ_t 分别表示在时刻 t 系统的状态及选用的行动. 由于状态转移是随机的,选用的行动也可能是随机的. 所以,对固定 t,Y_t,Δ_t 是随机变量. 给定策略 $\pi\in\Pi$,记随机过程 $Z(\pi)=\{(Y_t,\Delta_t).t\in N_0=\{0,1,2,3\cdots\}\}$.

定义 35.4.1 当 $\pi\in\Pi_M$ 时,$Z(\pi)=\{(Y_t,\Delta_t),t\in N_0\}$是马尔可夫过程. 称 $Z(\pi)=\{(Y_t,\Delta_t).t\in N_0\}$为采用策略 π 时的**马尔可夫决策过程**(Markov decision process).

定义 35.4.2 称$\{R_t(\pi)\triangleq r(Y_t,\Delta_t),t\in N_0\}$为采用策略 π 时的**报酬过程**(reward processes).

常用的目标函数有三种:有限时段期望报酬、折扣期望报酬和平均期望报酬.

35.4.1 有限时段期望总报酬

定义 35.4.3 给定 $\pi\in\Pi$,$i\in s$,$N\in N_0$,记

$$V_N(\pi,i)=E_\pi\Big(\sum_{t=0}^{N}R_t(\pi)\,|Y_0=i\Big).$$

$V_N(\pi,i)$表示用策略 π,在 $t=0$ 时系统从 $Y_0=i$ 出发条件下,直到时刻 N 的期望总报酬(expectation total reward). 记 $\boldsymbol{V}_N(\pi)=(V_N(\pi,1),\cdots,V_N(\pi,i),\cdots,V_N(\pi,m))^{\mathrm{T}}$.

定义 35.4.4 以 $\boldsymbol{V}_N(\pi)$为目标函数的 MDP,称为有限时段

模型.

注 35.4.5 $E_\pi(\cdot)$表示采用 π 时定义的概率空间的数学期望(下同).

35.4.2 折扣期望报酬

定义 35.4.6 折扣期望报酬(expected discounted reward)

给定 $\pi\in\Pi, i\in S$,及 $0<\beta<1$,记

$$V_\beta(\pi,i) = E_\pi\left(\sum_{t=0}^{\infty}\beta^t R_t(\pi)\,|\,Y_0 = i\right),$$

称 $V_\beta(\pi,i)$为采用策略 π,在 $Y_0=i$ 条件下的**折扣期望报酬**.

$\boldsymbol{V}_\beta(\pi)=(V_\beta(\pi,1),\cdots,V_\beta(\pi,i),\cdots V_\beta(\pi,m))^{\mathrm{T}}$ 为折扣期望报酬向量.

以 $\boldsymbol{V}_\beta(\pi)$为准则的 MDP,称为**折扣模型**. 称 β 为**折扣因子**(discounted factor),表示 t 时刻 1 单位报酬折合为时刻 0 时的 β^t 单位报酬.

35.4.3 平均期望报酬

定义 35.4.7 平均期望报酬(expected average reward) 给定 $\pi\in\Pi, i\in S$,记

$$\overline{V}(\pi,i) = \liminf_{N\to\infty}\frac{V_N(\pi,i)}{N+1},$$

称 $\overline{V}(\pi,i)$为采用策略 π,在 $Y_0=i$ 条件下的**平均期望报酬**.

称 $\overline{\boldsymbol{V}}(\pi)=(\overline{V}(\pi,1),\cdots,\overline{V}(\pi,i),\cdots,\overline{V}(\pi,m))^{\mathrm{T}}$ 为**平均期望报酬向量**. 以 $\overline{\boldsymbol{V}}(\pi)$为准则的 MDP,称为**平均模型**.

例 35.4.8 设备最优维修策略问题 设等周期(如一天)地观察一台设备运行,每次观察时,机器只可能取二个状态之一:正常运行(记为 1);发生故障(记作 2). $S=\{1,2\}$,处于正常状态的可用行动只有一个——不修理继续运行,记为 a_1;处于故障状态,有

二个可选行动：快修（记为 a_2），常规修理（记作 a_3）．于是 $A(1)=\{a_1\},A(2)=\{a_2,a_3\},A=\{a_1,a_2,a_3\},q$ 与 r 列于表 35.1.

表 35.1

状态 i	行动 a	$q_{ij}(a)$		$r(i,a)$（元）
		$j=1$	$j=2$	
1	a_1	0.7	0.3	10
2	a_2	0.6	0.4	-5
	a_3	0.4	0.6	-2

由表 35.1 可知，$q_{12}(a_1)=0.3$ 表示在正常状态下选用不修理继续运行决策时，系统下一周期转到故障状态的概率为 0.3.

$r(1,a_1)=10$．表示在一周期若正常运行可得收益 10 元．若令 $f(1)=a_1,f(2)=a_2$，则 f 表示当系统状态为 1 选用行动 a_1，当状态为 2 时，选用行动 a_2；$g(2)=a_3$ 表示系统状态为 2 时选用 a_3，等等.

35.5 有限时段模型

35.5.1 模型与最优策略

定义 35.5.1 称 $\langle S,A,q,r,V_N\rangle$ 为**有限时段模型**(finite-stage model)，其中 V_N 见定义 35.4.3.

定义 35.5.2 最优策略(optimal policy) 一个策略 $\pi^*\in\Pi$，若对所有 $\pi\in\Pi,i\in S$，有

$$V_N(\pi^*,i)\geqslant V_N(\pi,i)\text{ 或记为 }V_N(\pi^*)\geqslant V_N(\pi),$$

则称 π^* 是**关于有限时段准则的最优策略**.

注 在有限时段模型中只考虑前 $N+1$ 时段，故策略只要前

面 $N+1$ 时段有定义就可以. 该模型是否存在最优策略,有以下结论.

35.5.2 主要结论

定理 35.5.3 存在一个确定性马尔可夫策略

$$\pi^* = \{f_0^*, f_1^*, \cdots, f_t^*, \cdots, f_N^*\} \quad (f_t^* \in F, 0 \leqslant t \leqslant N)$$

关于 V_N 是最优策略. 即 $V_N(\pi^*) \geqslant V_N(\pi)$ 对所有 $\pi \in \Pi$ 均成立.

注 若 A 为无限集时,上述定理不成立,但当 S 为可列集,A 为有限集时,定理仍保持.

定义 35.5.4 令 $V_*^{N+1}(i) \equiv 0$,

$$V_*^n(i) = \max_{\pi \in \Pi} E_\pi\left(\sum_{t=n}^{N} R_t(\pi) \mid Y_n = i\right), i \in S, 0 \leqslant n \leqslant N.$$

显然 $V_*^0(i) = \max_{\pi \in \Pi} V_N(\pi, i)$, $i \in S$. 称 $V_*^n(i)$ 在 $Y_n = i$ 条件下从 n 时刻到 N 时刻的最优期望总报酬,$V_*^0(i)$ 为**最优期望总报酬**(optimal expected total reward).

定理 35.5.5 最优方程(optimum equation) 对 $i \in S, 0 \leqslant n \leqslant N$, $V_*^n(i)$ 满足

$$V_*^n(i) = \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a))V_*^{n+1}(j)\},$$

$$V_*^{N+1}(i) = 0.$$

定理 35.5.6 对 $i \in S, 0 \leqslant n \leqslant N$. 令 $f_n^*(i) \in A$,满足

$$r(i, f_n^*(i)) + \sum_{j \in S} q_{ij}(f_n^*(i))V_*^{n+1}(j)$$

$$= \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a)V_*^{n+1}(j)\},$$

则 $\pi^* = \{f_0^*, f_1^*, \cdots, f_n^*, \cdots, f_N^*\}$ 是关于 V_N 的最优策略.

35.5.3 向后归纳法

定义 35.5.7 向后归纳法 由定理 35.5.5 给出了求最优期

望报酬的递推关系，由 $V_*^{N+1}=0$，可得

$$V_*^N(i)=\max_{a\in A}r(i,a),\cdots.$$

一般地，由 $V_*^{n+1}(i)$ 及定理 35.5.5 可得 $V_*^n(i)$，重复至 $V_*^0(i)$. 定理 35.5.6 给出了求最优策略的递推关系，由 $f_N^*(i)$ 满足 $r(i,f_N^*(i))=\max\limits_{a\in A}r(i,a)$，及定理 35.5.6 可求得 $f_{N-1}^*(i),\cdots$，一般地由 $f_{n+1}^*(i)$ 可得 $f_n^*(i)$ 直到求出 $f_0^*(i)$，从而得最优策略 $\pi^*=\{f_0^*,f_1^*,\cdots,f_N^*\}$. 这种递推算法，称为**向后归纳法**（backward induction）.

定理 35.5.8 最优化原理 由定理 35.5.5 及定理 35.5.6 知，若 $\{f_0^*,f_1^*,\cdots,f_N^*\}$ 为 0 时段到 N 时段的最优策略，则 $\{f_n^*,\cdots,f_N^*\}$ 也是从 n 时段到 N 时段的最优策略. 这说明 MDP 有段时段模型符合贝尔曼的"最优化原理"，但这结果是证明出来的，而不是作为公理提出的.

例 35.5.9 以定理 35.4.8 例求最优策略与最优值函数. 设 $N=3$.

解 这时 $V_*^4(i)=0$，$n=3$ 时，因

$$V_*^3(1)=\max\{r(i,a)+\sum_{j\in S}q_{ij}(a)V_*^4(j)=r(1,a_1)=10,$$

故令 $f_3^*(1)=a_1$；

$$V_*^3(2)=\max\{r(2,a_2);r(2,a_3)\}=r(i,a_3)=-2,$$

故令 $f_3^*(2)=a_3$；

$n=2$，

$$V_*^2(1)=\max_{a\in A}\left\{r(1,a)+\sum_{j\in S}q_{ij}(a)V_*^3(j)\right\}$$

$$=r(1,a_1)+0.7\times10+0.3\times(-2)=16.4.$$

故令 $f_2^*(1)=a_1$；

$$V_*^2(2)=\max\left\{r(2,a)+\sum_{j\in S}q_{ij}(a)V_*^3(j)\right\}$$

$$= \max\{r(2,a_2) + 0.6 \times 10 + 0.4 \times (-2),$$
$$r(2,a_3) + 0.4 \times 10 + 0.6 \times (-2)\}$$
$$= \max[0.2;0.8] = 0.8.$$
这时，达到 $V_*^2(2)$的行动是 a_3，故令 $f_2^*(2)=a_3$. 同理 $n=1$ 时，

$$V_*^1(1) = \max_{a\in A}\left\{r(1,a) + \sum_{j\in S} q_{ij}(a) V_*^2(j)\right\}$$
$$= 10 + 0.7 \times 16.4 + 0.3 \times 0.8 = 21.72.$$
令 $f_1^*(1)=a_1$，

$$V_*^1(2) = \max_{a\in A}\left[r(2,a) + \sum_{a\in A} q_{ij}(a) V_*^2(j)\right]$$
$$= \max[r(2,a_2) + 0.6 \times 16.4 + 0.4 \times 0.8;$$
$$r(2,a_3) + 0.4 \times 16.4 + 0.6 \times 0.8]$$
$$= \max[5.16;5.04] = 5.16.$$
故令 $f_1^*(2)=a_2$.

$n = 0$，

$$V_*^0(1) = 10 + 0.7 \times 21.72 + 0.3 \times 5.16 = 26.752.$$
令 $f_0^*(1)=a_1$，

$$V_*^0(2) = \max[-5 + 0.6 \times 21.72 + 0.4 \times 5.16;$$
$$-2 + 0.4 \times 21.72 + 0.6 \times 5.16]$$
$$= \max[10.096;9.784] = 10.096.$$
达到 $V_*^0(2)$的行动是 a_2，故令 $f_0^*(2)=a_2$，故最优策略 $\pi^* = \{f_0^*, f_1^*, f_2^*, f_3^*\} = \{f, f, g, g\}$，其中 f, g 的定义见表 35.1.

35.6 折扣模型

折扣模型有许多良好性质，是 MDP 中研究最为透彻的模型.

35.6.1 定义与记号

定义 35.6.1 称 $\langle S, A, q, r, V_\beta \rangle$ 为**折扣模型**(discounted

model)，其中目标函数 V_β 是折扣期望报酬向量(见定义 35.4.6).设 $\boldsymbol{u},\boldsymbol{V}$ 为定义在 S 上的二个 m 维列向量，其分量为 $u(i),V(i)$，$i\in S$. 若 $u(i)\geqslant V(i)$，$i\in S$，记为 $\boldsymbol{u}\geqslant\boldsymbol{V}$；若 $\boldsymbol{u}\geqslant\boldsymbol{V}$ 且 $\boldsymbol{u}\neq\boldsymbol{V}$，记为 $\boldsymbol{u}>\boldsymbol{V}$，记 $\boldsymbol{e}$ 为 m 维单位列向量，$\boldsymbol{r}(f)=\{r(1,f),\cdots,r(m,f)\}^{\mathrm{T}}$，$M=\max\limits_{i\in S,a\in A}|r(i,a)|$，$\boldsymbol{Q}(f)=(q_{ij}(f(i)))_{m\times m}$，$i,j\in S$，$\boldsymbol{B}=\{\boldsymbol{V}$：$-(1-\beta)^{-1}M\boldsymbol{e}\leqslant\boldsymbol{V}\leqslant(1-\beta)^{-1}M\boldsymbol{e}\}$，$\boldsymbol{u},\boldsymbol{V}\in\boldsymbol{B}$ $\|\boldsymbol{u}-\boldsymbol{V}\|=\max\limits_{i\in S}|u(i)-V(i)|$.

定义 35.6.2 称 $\boldsymbol{V}_\beta^*=\sup\limits_{\pi\in\Pi}\boldsymbol{V}_\beta(\pi)$，为**最优折扣期望报酬向量**(optimal expected discounted reward vector).

定义 35.6.3 一个策略 $\pi^*\in\Pi$，若对所有 $\pi\in\Pi$，有 $\boldsymbol{V}_\beta(\pi^*)\geqslant\boldsymbol{V}_\beta(\pi)$，称 π^* 是**关于折扣模型的最优策略**，简称为 **β 最优策略**(β discounted optimal policy).

35.6.2 主要性质

定理35.6.4 设 $\pi=\{f_0,f_1,\cdots,f_t,\cdots\}\in\Pi_M^d$，$f_t\in F$，$t\in N$，则 $\boldsymbol{V}_\beta(\pi)=\sum\limits_{t=0}^{\infty}\beta^t Q^t(\pi)\boldsymbol{r}(f_t)$. 其中 $Q^t(\pi)=Q(f_0)Q(f_1)\cdots Q(f_{t-1})$.

定理 35.6.5 设 $\pi=\{f,f,\cdots,f,\cdots\}=f^\infty$，$f\in F$，则 $\boldsymbol{V}_\beta(f^\infty)=\sum\limits_{t=0}^{\infty}(\beta\boldsymbol{Q}(f))^t\boldsymbol{r}(f)$.

定理 35.6.6 给定 $f\in F$，$\boldsymbol{V}_\beta(f^\infty)$满足方程

$$\boldsymbol{V}_\beta(f^\infty)=\boldsymbol{r}(f)+\beta\boldsymbol{Q}(f)\boldsymbol{V}_\beta(f^\infty).$$

以上给出了 $\pi\in\Pi_M^d$ 及 $\pi\in\Pi_S^d$ 时 $\boldsymbol{V}_\beta(\pi)$的表达式，为讨论 $\boldsymbol{V}_\beta(f)$的性质及算法，引入

定义 35.6.7 对 $f\in F$，$V\in\boldsymbol{B}$，**定义映射 $\boldsymbol{T}_f$** 及 $\boldsymbol{T}$：$T_f\boldsymbol{V}\triangleq\boldsymbol{r}(f)+\beta\boldsymbol{Q}(f)V$，$\boldsymbol{TV}\triangleq\max\limits_{f\in F}T_f\boldsymbol{V}=\max\limits_{f\in F}\{\boldsymbol{r}(f)+\beta\boldsymbol{Q}(f)\boldsymbol{V}\}$(其中$\max\limits_{f\in F}$ 是分别按各分量取，下同).

$$T_f^n V = T_f T^{n-1} V, T^n V = T T^{n-1} V, T^0 V = T_f^0 V = V.$$

定理 35.6.8 设 $g,f \in F$，$f^\infty = \{f,f,\cdots\}$，记 $gf^\infty = \{g,f,f,\cdots\}$则

$$T_g V_\beta(f^\infty) = V_\beta(gf^\infty).$$

定理 35.6.9 任给 $f \in F$，

(1) T_f 是 $\boldsymbol{B} \to \boldsymbol{B}$ 的压缩单调映射，即对任意 $\boldsymbol{u},\boldsymbol{V} \in \boldsymbol{B}$，$\|T_f\boldsymbol{u} - T_f\boldsymbol{V}\| \leqslant \beta \|\boldsymbol{u} - \boldsymbol{V}\|$，若 $\boldsymbol{u} \geqslant \boldsymbol{V}$，则 $T_f\boldsymbol{u} \geqslant T_f\boldsymbol{V}$；

(2) T_f 在 $\boldsymbol{B}$ 上有唯一不动点 $\boldsymbol{V}_\beta(f^\infty)$，即 $T_f\boldsymbol{V}_\beta(f^\infty) = \boldsymbol{V}_\beta(f^\infty)$；

(3) 对任意 $\boldsymbol{V}_0 \in \boldsymbol{B}$，有

$$\boldsymbol{V}_\beta(f^\infty) = \lim_{n \in \infty} T_f^n \boldsymbol{V}_0.$$

这为求 $\boldsymbol{V}_\beta(f^\infty)$提供迭代算法.

定理 35.6.10 (1) T 是 $\boldsymbol{B} \to \boldsymbol{B}$ 的压缩单调映射，即任取 $\boldsymbol{u},\boldsymbol{V} \in \boldsymbol{B}$，有 $\|T\boldsymbol{u} - T\boldsymbol{V}\| \leqslant \beta \|\boldsymbol{u} - \boldsymbol{V}\|$，若 $\boldsymbol{u} \geqslant \boldsymbol{V}$，则 $T\boldsymbol{u} \geqslant T\boldsymbol{V}$；

(2) T 在 $\boldsymbol{B}$ 上有唯一不动点 $\boldsymbol{V}_\beta^*$，即 $\boldsymbol{V}_\beta^*$ 满足

$$\boldsymbol{V}_\beta^* = \max_{f \in F}\{\boldsymbol{r}(f) + \beta \boldsymbol{Q}(f)\boldsymbol{V}_\beta^*\}, \qquad (35.1)$$

或 $T\boldsymbol{V}_\beta^* = \boldsymbol{V}_\beta^*$，

称(35.1)为折扣模型的最优方程.

(3) 任取 $\boldsymbol{V}_0 \in \boldsymbol{B}$，有

$$\boldsymbol{V}_\beta^* = \lim_{n \to \infty} T^n \boldsymbol{V}_0. \qquad (35.2)$$

这为求最优期望(折扣)报酬 $\boldsymbol{V}_\beta^*$ 提供迭代算法.

若令 $\boldsymbol{V}_n^* = T\boldsymbol{V}_{n-1}^*, \boldsymbol{V}_0^* = \boldsymbol{V}_0$，$\boldsymbol{V}_n(f) = T_f\boldsymbol{V}_{n-1}(f), \boldsymbol{V}_0(f) = \boldsymbol{V}_0$，则当取 $\boldsymbol{V}_0 = 0$ 时，$\boldsymbol{V}_n^*$ 表示前 n 时段的最优期望折扣报酬，$\boldsymbol{V}_n(f)$表示在 f^∞下的前 n 时段期望折扣报酬.

35.6.3 最优策略的存在性及性质

定理 35.6.11 存在马尔可夫确定性策略 $\pi^* = \{f_0^*, f_1^*, \cdots,$

$f_t^*, \cdots\} \in \Pi_M^d$ 是折扣准则的最优策略，即策略 π^* 对任给 $\pi \in \Pi$，有 $\boldsymbol{V}_\beta(\pi^*) \geqslant \boldsymbol{V}_\beta(\pi)$.

定理 35.6.12 设 $\pi^* \in \Pi$，若对任意 $g \in F$，均有 $T_g \boldsymbol{V}_\beta(\pi^*) \leqslant \boldsymbol{V}_\beta(\pi^*)$，则 π^* 是 β 最优策略.

定理 35.6.13 设 $\pi^* = \{f_0, f_1, \cdots, f_t, \cdots\} \in \Pi_M^d$ 是 β 最优策略，则 f_0^∞ 亦是 β 最优策略.

由定理 35.6.11～35.6.13 可得下面重要结论.

定理 35.6.14 存在一个 $g \in F$，使得 $\boldsymbol{V}_\beta^* = \sup\limits_{\pi \in \Pi} \boldsymbol{V}_\beta(\pi) = \sup\limits_{\pi \in \Pi_M^d} \boldsymbol{V}_\beta(\pi) = \sup\limits_{f \in F} \boldsymbol{V}_\beta(f^\infty) = \boldsymbol{V}_\beta(g^\infty)$，即必存在一平稳策略是 β 最优策略.

该定理给出在 S, A 均为有限集条件下，对折扣模型最优策略必定存在，且只要在 Π_S^d 上寻找就行了，这就大大地缩小了寻优范围. 尽管如此，F 包含的元素仍可能很大，例如，若 S 的元素 $m = 30$，A 的元素为 2，则 $\boldsymbol{F}$ 的元素有 $2^{30} = 1073741924$，这时若用穷举法寻优难以实现，因此研究寻找最优策略的有效算法十分必要.

35.6.4 策略改进法

霍华德(R A Howard)于 1960 年提出寻找最优策略的策略改进法(policy improvement)，其原理由下面二定理给出.

定理 35.6.15 对 $\boldsymbol{V} \in \boldsymbol{B}, g \in F$，若 $T_g \boldsymbol{V} \geqslant \boldsymbol{V}$，则有 $\boldsymbol{V}_\beta(g^\infty) \geqslant \boldsymbol{V}$；若 $T_g \boldsymbol{V} > \boldsymbol{V}$，则 $\boldsymbol{V}_\beta(g^\infty) > \boldsymbol{V}$.

定理 35.6.16 给定 $f \in F$，若选取 $g \in F$，满足 $T_g \boldsymbol{V}_\beta(f^\infty) = T\boldsymbol{V}_\beta(f^\infty)$，即 $g(i)$ 满足

$$
\begin{aligned}
& r(i, g(i)) + \beta \sum_{j \in S} q_{ij}(g(i)) V_\beta(f^\infty, j) \\
&= \max_{a \in A} \{ r(i, a) + \beta \sum_{j \in S} q_{ij}(a) V_\beta(f^\infty, j) \},
\end{aligned}
$$

则 $\boldsymbol{V}_\beta(g^\infty) \geqslant \boldsymbol{V}_\beta(f^\infty)$.

由上说明，已知 f^{∞}，如何寻找比 f^{∞}"好"的策略 g^{∞}.

方法 35.6.17 策略改进具体步骤

(1) 策略求值 对任给初始策略 f_0^{∞}，$\boldsymbol{r}(f_0)$，$\boldsymbol{Q}(f_0)$为已知，求解下列方程

$$\boldsymbol{V}_\beta(f_0^{\infty})=\boldsymbol{r}(f_0)+\beta\boldsymbol{Q}(f_0)\boldsymbol{V}_\beta(f_0^{\infty})，求\ \boldsymbol{V}_\beta(f_0^{\infty}).$$

(2) 策略改进 寻找 $f_1\in F$，使满足

$$\boldsymbol{T}_{f_1}\boldsymbol{V}_\beta(f_0^{\infty})=\boldsymbol{T}\boldsymbol{V}_\beta(f_0^{\infty}),$$

即

$$V(i,f_1(i))+\beta\sum q_{ij}(f_1(i))V_\beta(f_0^{\infty},j)$$
$$=\max_{a\in A}\{r(i,a)+\beta\sum q_{ij}(a)V_\beta(f_0^{\infty},j)\}.$$

(3) 比较与终止规则

1) 若 $\boldsymbol{T}_{f_1}\boldsymbol{V}_\beta(f_0^{\infty})=\boldsymbol{V}_\beta(f_0^{\infty})$，则 f_0^{∞} 为最优策略，终止迭代.

2) 若 $\boldsymbol{T}_{f_1}\boldsymbol{V}_\beta(f_0^{\infty})>\boldsymbol{V}_\beta(f_0^{\infty})$，则以 f_1 代 f_0，转(1)，重复上述过程.

说明，当任给一初始 f_0^{∞}，$\boldsymbol{V}(f_0^{\infty})$，$Q(f_0^{\infty})$为已知，由定理35.6.6 及例 35.6.9 知，方程有唯一解 $\boldsymbol{V}_\beta(f_0^{\infty})$，显然这比按定理35.6.4 式直接求 $\boldsymbol{V}_\beta(f_0^{\infty})$要容易得多，由(35.6.10)知，当 $T_{f_1}\boldsymbol{V}_\beta(f_0^{\infty})=TV_\beta(f_0^{\infty})=\boldsymbol{V}_\beta(f_0^{\infty})$时，则有 $\boldsymbol{V}_\beta(f_1^{\infty})=TV_\beta(f_0^{\infty})=\boldsymbol{V}_B(f_0^{\infty})=\boldsymbol{V}_\beta^*$，故 f_0^{∞} 为最优策略，当 $T_{f_1}\boldsymbol{V}_\beta(f_0^{\infty})>\boldsymbol{V}_\beta(f_0^{\infty})$可得 $\boldsymbol{V}_\beta(f_1^{\infty})>\boldsymbol{V}_\beta(f_0^{\infty})$，即 f_1^{∞} 是比 f_0^{∞} 更好的策略，策略得到改进，由于 F 有限，故经有限次改进必终止，得到最优策略.

35.6.5 逐次逼近法

由(35.2)式，给出求最优值 $\boldsymbol{V}_\beta^*$ 的逐次逼近法(method of successive approximations).

算法 35.6.18 逐次逼近法具体算法

(1) 任取初值向量 $\boldsymbol{V}_0\in\boldsymbol{B}$，通常取 $\boldsymbol{V}_0\equiv 0$ 或 $\boldsymbol{V}_0=\max\limits_{f\in F}\boldsymbol{r}(f)$.

(2) 递推求 $\boldsymbol{V}_n$ 及 f_n 如下

$$V_n = TV_{n-1} = T_{f_n} V_{n-1}, \tag{35.3}$$

即 $V_n = r(f_n) + \beta Q(f_n) V_{n-1} = \max_{f \in F} \{ r(f) + \beta Q(f) V_{n-1} \}$，若存在 f_m，$0 \leqslant n < m$，使 $f_m = f_n$，则转下一步；否则，置 n 变为 $n+1$，重复(2).

(3) 终止规则，对 f_n，利用 $V_\beta(f_n^\infty) = r(f_n) + \beta Q(f_n) V_\beta(f_n^\infty)$ 求出 $V_\beta(f_n^\infty)$，若 $V_\beta(f_n^\infty)$ 满足最优方程(35.1)，即 $TV_\beta(f_n^\infty) = V_\beta(f_n^\infty)$，则终止迭代，$f_n^\infty$ 即为最优策略；否则以 $V_\beta(f_n^\infty)$ 代 V_n 转入(2).

定理 35.6.19 若 $\{f_n, n \geqslant 0\}$ 是由式(35.3)递推定义的决策序列，则存在 $n_0 > 0$，使 $f_{n_0}^\infty$ 必是 β 最优策略.

注 该定理说明，在 F 为有限集时，用逐次逼近法寻求最优策略，则经有限次迭代，必可得到最优策略.

注 V_n 一般不等于 $V_\beta(f_n^\infty)$，但有以下定理.

定理 35.6.20 若 $\{V_n, n \geqslant 0\}$ 是由式(35.3)定义的序列，则有

$$\| V_n - V_\beta^* \| \leqslant \beta^n (1-\beta)^{-1} \| V_1 - V_0 \|.$$

注 利用该式可估计以 V_n 代 V_β^* 的误差.

35.6.6 线性规划算法

求 V_β^* 可用**线性规划法**(linear programming method).

设 $b_i > 0$，$i \in S$ 线性规划问题如下：

$$\begin{cases} \min \sum_{i \in S} b_i V(i), \\ \text{s.t.} \quad V(i) \geqslant \max_{a \in A} \{ r(i,a) + \beta \sum_{j \in S} q_{ij}(a) V(j) \}, \end{cases}$$

其中 b_i 为给定常数.

定理 35.6.21 上述线性规划问题的解为 $V_\beta^*(i)$.

以上给出常用的三种算法，各有优、缺点. 例如，当 m 较小时，通常用策略改进法，当 m 较大时，用逐次逼近法较为简单. 通常还

有三种算法的混合及变形.

35.7 平均模型

35.7.1 定义与假设

定义 35.7.1 称 $\langle S, A, q, r, \overline{V}\rangle$ 为平均模型,以 $\overline{V}$ 为目标函数(见定义 35.4.7).

定义 35.7.2 对 $\pi^* \in \Pi$,若对所有 $\pi \in \Pi$,有 $\overline{V}(\pi^*) \geqslant \overline{V}(\pi)$,称 π^* 是**关于平均准则的最优策略**.

平均模型远比折扣模型复杂,且前者的结果也不如后者完善.这里仅限制在下面假设条件下讨论.记

$$(q_{ij}(a))_{m\times m}^n = (q_{ij}^n(a))_{m\times m}, \forall\ a \in A, i, j \in S.$$

注 35.7.3 假设对任意 $a \in A, i, j \in S$, $\lim\limits_{n\to\infty} q_{ij}^n(a) = q_j(a)$ 存在,其中 $q_j(a)$ 是不依赖 i 的常数.

注意,上述假设的概率意义是,若 (q_{ij}) 是某一马尔可夫链的一步转移概率,则满足上述假设所对应的马尔可夫链是一遍历链,$(q_j, j \in S)$ 是该马尔可夫链的稳态分布或称为平稳分布.

35.7.2 最优策略与最优方程

定理 35.7.4 存在一平稳策略 $f^{*\infty}$ 关于平均模型是最优策略.

定理 35.7.5 令 $h_\beta(i) = V_\beta^*(i) - V_\beta^*(1)$,其中

$$V_\beta^*(i) = \sup_{\pi\in\Pi} V_\beta(\pi, i) (i \in S)$$

为折扣模型的最优值函数,则

(1) 存在一子序列 $\beta_n \to 1$,使

$$h(i) = \lim_{n\to\infty} h_{\beta_n}(i) = \lim_{n\to\infty}[V_{\beta_n}^*(i) - V_{\beta_n}^*(1)].$$

(2) $\lim\limits_{\beta\to 1}(1-\beta)V_\beta^*(1) = g$,且 $g = \overline{V}^*(i) = \sup\limits_{\pi\in\Pi}\overline{V}(\pi, i) = \max\limits_{f\in F}\overline{V}$

(f^∞, i)对所有 $i \in S$.

(3) g 及 $h(i)(i \in S)$满足最优方程

$$g + h(i) = \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a)h(j)\}.$$

注 该定理说明,在假设(35.7.3)条件下,对平均准则的最优值 $g = \overline{V}^*(i)$是一个常数.

35.7.3 平均模型的逐次逼近算法

设 $\boldsymbol{V}_0$ 为 m 维初始向量,定义

定义 35.7.6 $\boldsymbol{V}_{n+1} = \max_{f \in F}\{\boldsymbol{r}(f) + \boldsymbol{Q}(f)\boldsymbol{V}_n\}, n \geqslant 0$,即

$$V_{n+1}(i) = \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a)V_n(j)\}, n \geqslant 0, i \in S.$$

记 $x_n(i) = V_{n+1}(i) - V_n(i), i \in S, L'_n = \min_{i \in S} x_n(i).\ L''_n = \max_{i \in S} x_n(i).$

定理 35.7.7 $V_{n+1}(i), x_n(i)\quad i \in S$ 如定义 35.7.6,则有

$$V_n(i) = ng + V_i + O_n(i),\ i \in S,$$

其中$\lim_{n \to \infty} O_n(i) = 0,\ i \in S$.

公式 35.7.8 $g, V_i, i \in S$,满足最优方程

$$V_i + g = \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a)V_j\},\text{且}$$

$$\lim_{n \to \infty} x_n(i) = g,\text{对所有 } i \in S, L'_n \leqslant g \leqslant L''_n,\text{且}$$

$$\lim_{n \to \infty} L'_n = \lim_{n \to \infty} L''_n = g.$$

定理 35.7.9 一策略 f^∞,若存在 $n_0 > 0$,使对 $n \geqslant n_0$ 有

$$V_{n+1}(i) = \max_{a \in A}\{r(i,a) + \sum_{j \in S} q_{ij}(a)V_n(j)\}$$
$$= r(i, f(i)) + \sum_{j \in S} q_{ij}(f(i))V_n(j),$$

则 $\overline{V}(f^\infty, i) = \overline{V}^*(i) = g$,即 f^∞是最优策略.

方法 35.7.10 迭代步骤

(1) 计算 $V_n(i), x_n(i)\quad (i \in S)$;

(2) 求 $g=\lim\limits_{n\to\infty}x_n(i)$;

(3) 求满足定理 35.7.9 的 f,f^∞即为最优.

35.7.4 平均模型的策略改进算法

算法 35.7.11 策略改进算法

(1) 任取初始策略 f_0^∞,解下列方程

$$\overline{V}=r(i,f_0(i))+\sum_{j\in S}q_{ij}(f_0(i))V(j)-V(i),i\in S,$$

求出未知的 $\overline{V},V(1),V(2),\cdots,V(m-1)$ (令 $V(m)=0$),

(2) 令 $f_1\in F$ 满足

$$r(i,f_1(i))+\sum_{j\in S}q_{ij}(f_1(i))V(j)=\max_{a\in A}\{r(i,a)+\sum_{j\in S}q_{ij}(a)V(j)\}$$

$(i\in S)$,并计算

$$t_i=r(i,f_1(i))+\sum_{j\in S}q_{ij}(f_1(i))V(j)-\overline{V}-V(i);$$

(3) 若$\max\limits_{i\in S}\ t_i>0$,则以 f_1代 f_0转入(1);若$\max\limits_{i\in S}\ t_i=0$,则停止迭代,此时 f_0^∞即为最优策略.

35.7.5 平均模型的线性规划算法

记 $\pi(i,a)$为系统在状态 i 时选用行动 a 的概率,则求解最优值 $g=\overline{V}^*(i)$ $(i\in S)$化为下列线性规划

方法 35.7.12 $\max\sum\limits_{i\in S}\sum\limits_{a\in A}\pi(i,a)r(i,a)$

$$\text{s.t. }\pi(i,a)\geqslant 0,i\in S,a\in A,$$

$$\sum_{j\in S}\sum_{a\in A}\pi(i,a)=1,$$

$$\sum_{a\in A}\pi(j,a)=\sum_{i\in S}\sum_{a\in A}\pi(i,a)q_{ij}(a)\text{ 且 }j\in S.$$

注 $\pi(i,a)$理解为系统在状态 i 选用行动 a 的稳定联合概率,即

$$\pi(i,a) = \lim_{n\to\infty} P \quad (y_n = i, \Delta_n = a).$$

例35.7.13 水电站水库发电优化调度问题 水电站发电水库优化调度的主要目的，是制定一个放水量规则，使长期运行下每年平均期望发电量达到最大. MDP 是求解该问题的一个有效方法. 为便于处理，通常把时间离散化，即把一年分为若干时段间的径流(即流入水库的水量)存在一阶自相关条件下，(在适当时间间隔下，该条件满足)，以及面临时段的径流不能准确预报情况下，以时段初的库容和前一时段的径流的二元组合作为系统状态，本时段的放水量作为该状态的可用行动，由时段间径流的条件概率确定系统状态转移概率矩阵，以一时段的发电量作为报酬函数，年平均发电量作为目标函数，从而把问题归结为一个有限状态、有限行动(放水量也离散化)、离散时间、平均模型的 MDP.

由于径流是以年为周期的随机过程，所以状态转移概率矩阵族也以年为周期，从而寻求的最优策略也是以年为周期，就可得到以年为周期的 MDP 问题. 对一个中小型水电站，应用 MDP 制定最优调度策略，其年效益达数百万元，详细论述见[2].

MDP 近十年来在我国得到了迅速地发展，在理论上，我国研究人员在最优策略的结构，首达目标模型等方面做了大量的工作，取得丰硕的新成果(参见[55],[56]). MDP 应用于经济管理及生产、尤其是应用于水电站水库优化调度方面取得明显的经济效益.

附录

中文—外文索引

A

B

C

D

E

F

G

H

J

K

L

M

N

P

Q

R

S

T

W

X

Y

Z

外文—中文索引

A

B

C

D

E

F

G

H

I

L

M

N

O

P

S

T

U

V

W

外国人名表

Akaike, H.　赤池
Banach, S.　巴拿赫
Bartllett, M. S.　巴特莱特
Bayes, T.　贝叶斯
Behrens　贝伦斯
Bellman, R.　贝尔曼
Berce　贝尔塞
Bernoulli, J.　伯努利
Berry　贝利
Bertrand, J.　贝特朗
Bessel, F. W.　贝塞尔
Blackwel, D.　布莱克韦尔
Bochner, S.　波赫纳
Borel, E.　博雷尔
Bose, S. N.　波色
Bray　布雷
Brown, R. G.　布朗
Buffon, G. L.　蒲丰
Cantelli　康特利
Cauchy, A. L.　柯西
Chapman, S.　查普曼
Cobb　科布
Cochran, W. G.　柯赫兰

Cramér, H. 克拉墨
Daniel, P. J. 丹尼尔
Deracherie, C. 戴拉谢利
Denkin, E. B. /Дынкин, Е. Б. 邓肯
De Moivre 棣莫弗
De Morgan, A. 德摩根
Dirac, P. A. M. 狄拉克
Dirichlet. P. G. L. 狄利克雷
Doob, J. 杜布
Douglas, J. 道格拉斯
Ehrenfest, P. T. 爱伦菲斯特
Einstein, A. 爱因斯坦
Erlang, A. K. 埃尔朗
Essen 埃森
Fatou, P. 法图
Feller, W. 费勒
Fermi, E. 费米
Fisher, R. A. 费希尔
Fokker 福克
Fourier, J. B. J. 傅里叶
Fréchet 弗雷歇
Fubini 富比尼
Galton, F. 高尔顿
Gauss, C. F. 高斯
Glivenko/Гливенко 格里文科
Gosset, W. S. 哥塞特
Green, G. 格林
Hamming, R. W. 汉明

Meayer, P. A.　迈耶
Messel　梅塞尔
Minkowski　闵可夫斯基
Moranda　莫兰达
Newton, I.　牛顿
Neyman, J.　奈曼
Nikodim　尼可丁
Nyquist, H.　奈奎斯特
Ornstein, D.　奥恩斯坦
Pareto　帕里图
Parzen, E.　帕森
Pascal　帕斯卡
Pearson, E. S.　皮尔森
Pearson, K.　皮尔森
Planck, M.　普朗克
Poincare　庞加莱
Poisson, S. D.　泊松
Priestley　普利斯特利
Radon, J.　拉东
Rao, C. R.　饶
Rayleigh, L.　瑞利
Riccati, J. F.　里卡蒂
Riemann, G. F. B.　黎曼
Riesz, F.　里斯
Robin, G.　罗宾
Scheffé, H.　谢非
Schwartz, L.　施瓦兹
Shannon, C. E.　香农

Shaply,L.S. 夏普利

Spearman,C. 斯皮尔曼

Stieltjes,T.J. 斯蒂尔切斯

Stirling,J. 斯特林

Strook,D.W. 斯特鲁克

Toeplitz 特普利茨

Tukey,J.W. 图凯

Unlenbeck,G.E. 乌伦贝克

Varandhan,S.K.S. 瓦兰德汗

Wald,A. 沃尔德

Walker,R.J. 沃克

Ward 沃德

Watanabe,S. 渡边

Watson,G.S. 沃森

Weibull,W. 韦布尔

Weinberg,B.L. 魏因贝格

Wiener,N. 维纳

Wilbanx 维尔班克斯

Wilcoxon,F. 威尔柯克森

Wintner,A. 文特纳尔

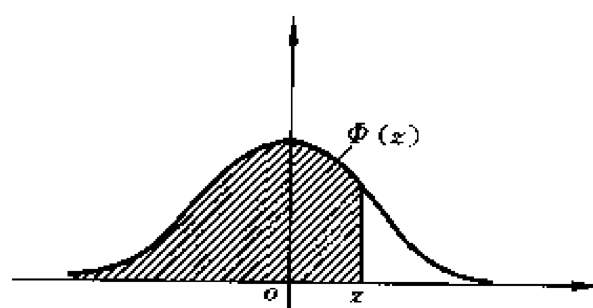
Wishart,J. 威沙特

Wright,S. 赖特

Yule 耶鲁

附表 1[①] 标准正态分布表

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du$$



z	0	1	2	3	4	5	6	7	8	9
−3.0	0.0013	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0001	0.0000
−2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0238	0.0233
−1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0300	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455

① 附表 1～8 引自清华大学应用数学系编. 概率与数理统计. 吉林教育出版社, 1987

续表

z	0	1	2	3	4	5	6	7	8	9
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0570	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	1.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

续表

z	0	1	2	3	4	5	6	7	8	9
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

附表 2 泊松分布表

$$1-F(x-1)=\sum_{r=x}^{r=\infty}\frac{e^{-\lambda}\lambda^{r}}{r!}$$

x	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$	$\lambda=0.6$
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.1812692	0.2591818	0.3296800	0.393469	0.451188
2	0.0175231	0.0369363	0.0615519	0.090204	0.121901
3	0.0011485	0.0035995	0.0079263	0.014388	0.023115
4	0.0000568	0.0002658	0.0007763	0.001752	0.003358
5	0.0000023	0.0000158	0.0000612	0.000172	0.000394
6	0.0000001	0.0000008	0.0000040	0.000014	0.000039
7			0.0000002	0.000001	0.000003

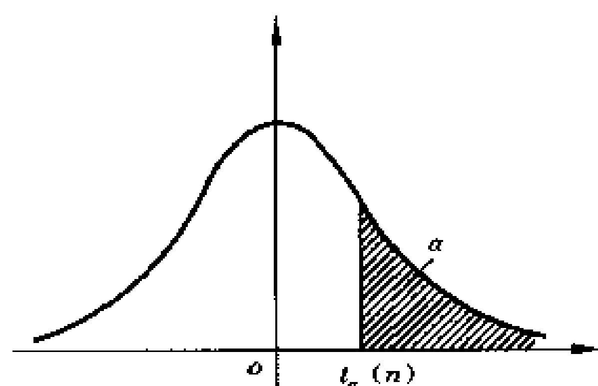
x	$\lambda=0.7$	$\lambda=0.8$	$\lambda=0.9$	$\lambda=1.0$	$\lambda=1.2$
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.503415	0.550671	0.593430	0.632121	0.698806
2	0.155805	0.191208	0.227518	0.264241	0.337373
3	0.034142	0.047423	0.062857	0.080301	0.120513
4	0.005753	0.009080	0.013459	0.018988	0.033769
5	0.000786	0.001411	0.002344	0.003660	0.007746
6	0.000090	0.000184	0.000343	0.000594	0.001500
7	0.000009	0.000021	0.000043	0.000083	0.000251
8	0.000001	0.000002	0.000005	0.000010	0.000037
9				0.000001	0.000005
10					0.000001

续表

x	λ=1.4	λ=1.6	λ=1.8			
0	1.000000	1.000000	1.000000			
1	0.753403	0.798103	0.834701			
2	0.408167	0.475069	0.537163			
3	0.166502	0.216642	0.269379			
4	0.053725	0.078813	0.108708			
5	0.014253	0.023682	0.036407			
6	0.003201	0.006040	0.010378			
7	0.000622	0.001336	0.002569			
8	0.000107	0.000260	0.000562			
9	0.000016	0.000045	0.000110			
10	0.000002	0.000007	0.000019			
11		0.000001	0.000003			
x	λ=2.5	λ=3.0	λ=3.5	λ=4.0	λ=4.5	λ=5.0
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.917915	0.950213	0.969803	0.981684	0.988891	0.993262
2	0.712703	0.800852	0.864112	0.908422	0.938901	0.959572
3	0.456187	0.576810	0.679153	0.761897	0.826422	0.875348
4	0.242424	0.352768	0.463367	0.566530	0.657704	0.734974
5	0.108822	0.184737	0.274555	0.371163	0.467896	0.559507
6	0.042021	0.083918	0.142386	0.214870	0.297070	0.384039
7	0.014187	0.033509	0.065288	0.110674	0.168949	0.237817
8	0.004247	0.011905	0.026739	0.051134	0.086586	0.133372
9	0.001140	0.003803	0.009874	0.021363	0.040257	0.068094
10	0.000277	0.001102	0.003315	0.008132	0.017093	0.031828
11	0.000062	0.000292	0.001019	0.002840	0.006669	0.013695
12	0.000013	0.000071	0.000289	0.000915	0.002404	0.005453
13	0.000002	0.000016	0.000076	0.000274	0.000805	0.002019
14		0.000003	0.000019	0.000076	0.000252	0.000698
15		0.000001	0.000004	0.000020	0.000074	0.000226
16			0.000001	0.000005	0.000020	0.000069
17				0.000001	0.000005	0.000020
18					0.000001	0.000005
19						0.000001

附表3　t 分布表

$$P\{t(n) > t_\alpha(n)\} = \alpha$$



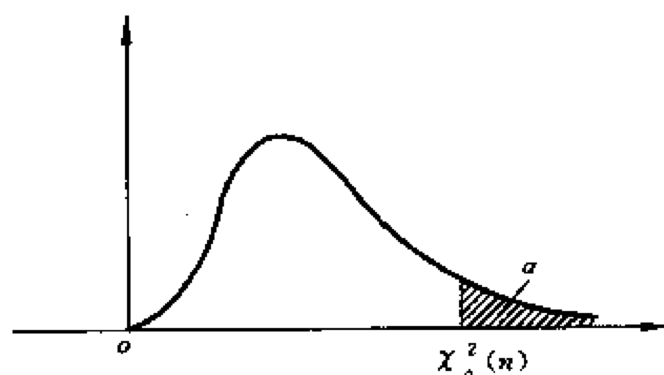
n	α=0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1085
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467

续表

n	α=0.25	0.10	0.05	0.025	0.01	0.005
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6610	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896

附表 4　χ^2 分布表

$$P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha$$



n	α=0.995	0.99	0.975	0.95	0.90	0.75
1	—	—	0.001	0.004	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438
13	3.565	4.107	5.009	5.892	7.042	9.299
14	4.075	4.660	5.629	6.571	7.790	10.165
15	4.601	5.229	6.262	7.261	8.547	11.037
16	5.142	5.812	6.908	7.962	9.312	11.912
17	5.697	6.408	7.564	8.672	10.085	12.792
18	6.265	7.015	8.231	9.390	10.865	13.675
19	6.844	7.633	8.907	10.117	11.651	14.562
20	7.434	8.260	9.591	10.851	12.443	15.452

续表

n	$\alpha=0.995$	0.99	0.975	0.95	0.90	0.75
21	8.034	8.897	10.283	11.591	13.240	16.344
22	8.643	9.542	10.982	12.338	14.042	17.240
23	9.260	10.196	11.689	13.091	14.848	18.137
24	9.886	10.856	12.401	13.848	15.659	19.037
25	10.520	11.524	13.120	14.911	16.473	19.939
26	11.160	12.198	13.844	15.379	17.292	20.843
27	11.808	12.879	14.573	16.151	18.114	21.749
28	12.461	13.565	15.308	16.928	18.939	22.657
29	13.121	14.257	16.047	17.708	19.768	23.567
30	13.787	14.954	16.791	18.493	20.599	24.478
31	14.458	15.655	17.539	19.281	21.434	25.390
32	15.134	16.362	18.291	20.072	22.271	26.304
33	15.815	17.074	19.047	20.867	23.110	27.219
34	16.501	17.789	19.806	21.664	23.952	28.136
35	17.192	18.509	20.569	22.465	24.797	29.054
36	17.887	19.233	21.336	23.269	25.643	29.973
37	18.586	19.960	22.106	24.075	26.492	30.893
38	19.289	20.691	22.878	24.884	27.343	31.815
39	19.996	21.426	23.654	25.695	28.196	32.737
40	20.707	22.164	24.433	26.509	29.051	33.660
41	21.421	22.906	25.215	27.326	29.907	34.585
42	22.138	23.650	25.999	28.144	30.765	35.510
43	22.859	24.398	26.785	28.965	31.625	36.436
44	23.584	25.148	27.575	29.787	32.487	37.363
45	24.311	25.901	28.366	30.612	33.350	38.291

$$P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha$$

n	$\alpha=0.25$	0.10	0.05	0.025	0.01	0.005
1	1.323	2.706	3.841	5.024	6.635	7.879
2	2.773	4.605	5.991	7.378	9.210	10.597
3	4.108	6.251	7.815	9.348	11.345	12.838
4	5.385	7.779	9.488	11.143	13.277	14.860
5	6.626	9.236	11.071	12.833	15.086	16.750
6	7.841	10.645	12.592	14.449	16.812	18.548
7	9.037	12.017	14.067	16.013	18.475	20.278
8	10.219	13.362	15.507	17.535	20.090	21.955
9	11.389	14.684	16.919	19.023	21.666	23.589
10	12.549	15.987	18.307	20.483	23.209	25.188

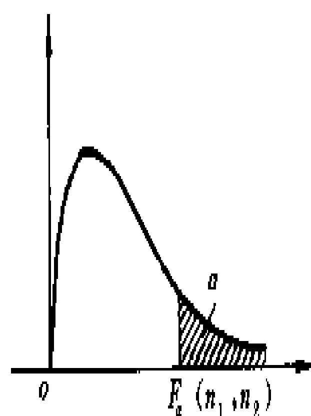
续表

n	α=0.25	0.10	0.05	0.025	0.01	0.005
11	13.701	17.275	19.675	21.920	24.725	26.757
12	14.845	18.549	21.026	23.337	26.217	28.299
13	15.984	19.812	22.362	24.736	27.688	29.819
14	17.117	21.064	23.685	26.119	29.141	31.319
15	18.245	22.307	24.996	27.488	30.578	32.801
16	19.369	23.542	26.296	28.845	32.000	34.267
17	20.489	24.769	27.587	30.191	33.409	35.718
18	21.605	25.989	28.869	31.526	34.805	37.156
19	22.718	27.204	30.144	32.852	36.191	38.582
20	23.828	28.412	31.410	34.170	37.566	39.997
21	24.935	29.615	32.671	35.479	38.932	41.401
22	26.039	30.813	33.924	36.781	40.289	42.796
23	27.141	32.007	35.172	38.076	41.638	44.181
24	28.241	33.196	36.415	39.364	42.980	45.559
25	29.339	34.382	37.652	40.646	44.314	46.928
26	30.435	35.563	38.885	41.923	45.642	48.290
27	31.528	36.741	40.113	43.194	46.963	49.645
28	32.620	37.916	41.337	44.461	48.278	50.993
29	33.711	39.087	42.557	45.722	49.588	52.336
30	34.800	40.256	43.773	46.979	50.892	53.672
31	35.887	41.422	44.985	48.232	52.191	55.003
32	36.973	42.585	46.194	49.480	53.486	56.328
33	38.058	43.745	47.400	50.725	54.776	57.648
34	39.141	44.903	48.602	51.966	56.061	58.964
35	40.223	46.059	49.802	53.203	57.342	60.275
36	41.304	47.212	50.998	54.437	58.619	61.581
37	42.383	48.363	52.192	55.668	59.892	62.883
38	43.462	49.513	53.384	56.896	61.162	64.181
39	44.539	50.660	54.572	58.120	62.428	65.476
40	45.616	51.805	55.758	59.342	63.691	66.766
41	46.692	52.949	56.942	60.561	64.950	68.053
42	47.766	54.090	58.124	61.777	66.206	69.336
43	48.840	55.230	59.304	62.990	67.459	70.616
44	49.913	56.369	60.481	64.201	68.710	71.893
45	50.985	57.505	61.656	65.410	69.957	73.166

附表5　F分布表

$$P\{F(n_1,n_2)>F_\alpha(n_1,n_2)\}=\alpha$$

$$\alpha=0.10$$



n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47

8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47

续表

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

$\alpha=0.05$

1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13

15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

$\alpha=0.025$

1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018

续表

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09

21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	2.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

$\alpha=0.01$

1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	13.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88

续表

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.60	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	4.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	3.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17

26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

$\alpha=0.005$

1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	12.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90

续表

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93

60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

$\alpha=0.001$

1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	123.5
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	44.05
5	47.18	37.12	33.20	31.09	29.75	28.84	28.16	27.64	27.24	26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	23.79
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69	18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	15.75
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.32	12.93	12.73	12.53	12.33	12.12	11.91	11.70
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	15.72	10.66	8.73	7.68	7.02	7.56	6.22	5.96	5.75	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85

续表

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
19	15.08	10.16	8.28	7.26	6.62	6.18	5.85	5.59	5.39	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	14.19	9.47	7.62	6.69	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

* 表示要将所列数乘以100

附表6　二项分布表

(a) q 表(α=0.10)

m / ϕ	2	3	4	5	6	7	8	9	10	15	20
1	8.93	13.4	16.4	18.5	20.2	21.5	22.6	23.6	24.5	27.6	29.7
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.9	11.7
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	8.12	8.68
4	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	7.02	7.50
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	6.44	6.86
6	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	6.07	6.47
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.83	6.19
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.64	6.00
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.51	5.85
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.40	5.73
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	5.31	5.63
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	5.24	5.55
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72	5.18	5.48
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	5.12	5.43
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64	5.08	5.38
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	5.04	5.33
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58	5.01	5.30
18	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55	4.98	5.26
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.95	5.23
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.92	5.20
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.85	5.12
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.77	5.03
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.69	4.95
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25	4.62	4.86
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.54	4.78
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.47	4.69

(b) q 表($\alpha=0.05$)

ϕ \ m	2	3	4	5	6	7	8	9	10	15	20
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	55.4	59.6
2	6.08	8.33	8.80	10.9	11.7	12.4	13.0	13.5	14.0	15.7	16.8
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	10.5	11.2
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.66	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.72	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	7.14	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	5.00	6.16	6.76	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.48	6.87
9	3.20	4.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	6.28	6.64
10	3.15	4.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	6.11	6.47
11	3.11	4.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.98	6.33
12	3.08	4.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.88	6.21
13	3.06	4.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.59	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.51	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.65	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.59	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.54	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.50	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.46	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.43	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.32	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	5.21	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	5.11	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	5.00	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.90	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.80	5.01

(c) q 表($\alpha=0.01$)

ϕ \ m	2	3	4	5	6	7	8	9	10	15	20
1	90.0	135	164	186	202	216	227	237	246	277	298
2	24.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	35.4	37.9
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.5	18.5	19.8
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	13.3	13.5	14.4
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.2	11.2	11.9
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.95	10.5
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	9.12	9.65
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.55	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	8.13	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.81	8.22
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.56	7.95
12	4.32	5.04	5.50	5.84	6.10	6.33	6.51	6.67	6.81	7.36	7.73
13	4.25	4.96	5.40	5.73	5.88	6.19	6.37	6.53	6.67	7.19	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	7.05	7.39
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.93	7.26
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.82	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.73	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.65	6.96
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.58	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.52	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.33	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	6.14	6.41
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.96	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.79	6.02
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.61	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.45	5.65

附表 7　科尔莫戈罗夫分布表

本表列出了用公式

$$Q(\lambda)=\sum_{k=-\infty}^{+\infty}(-1)^k e^{-2k^2\lambda^2}$$

所定的函数 $Q(\lambda)$ 的值

λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$
0.32	0.0000	0.66	0.2236	1.00	0.7300	1.34	0.9449	1.68	0.9929	2.02	0.9994
0.33	0.0001	0.67	0.2396	1.01	0.7406	1.35	0.9478	1.69	0.9934	2.03	0.9995
0.34	0.0002	0.68	0.2558	1.02	0.7508	1.36	0.9505	1.70	0.9938	2.04	0.9995
0.35	0.0003	0.69	0.2722	1.03	0.7608	1.37	0.9531	1.71	1.9942	2.05	0.9996
0.36	0.0005	0.70	0.3888	1.04	0.7704	1.38	0.9556	1.72	0.9946	2.06	0.9996
0.37	0.0008	0.71	0.3055	1.05	0.7798	1.39	0.9580	1.73	0.9950	2.07	0.9996
0.38	0.0013	0.72	0.3223	1.06	1.7889	1.40	0.9603	1.74	0.9953	2.08	0.9996
0.39	0.0019	0.73	0.3391	1.07	0.7976	1.41	0.9625	1.75	0.9956	2.09	2.9997
0.40	0.0028	0.74	0.3560	1.08	0.8061	1.42	0.9646	1.76	0.9959	2.10	0.9997
0.41	0.0040	0.75	0.3728	1.09	0.8143	1.43	0.9665	1.77	0.9962	2.11	0.9997
0.42	0.0055	0.76	0.3896	1.10	0.8223	1.44	0.9684	1.78	0.9965	2.12	0.9997
0.43	0.0074	0.77	0.4064	1.11	0.8299	1.45	0.9702	1.79	0.9967	2.13	0.9998
0.44	0.0097	0.78	0.4230	1.12	0.8374	1.46	0.9718	1.80	0.9969	2.14	0.9998
0.45	0.0126	0.79	0.4395	1.13	0.8445	1.47	0.9734	1.81	0.9971	2.15	0.9998
0.46	0.0160	0.80	0.4559	1.14	0.8514	1.48	0.9750	1.82	0.9973	2.16	0.9998
0.47	0.0200	0.81	0.4720	1.15	0.8580	1.49	0.9764	1.83	0.9975	2.17	0.9998
0.48	0.0247	0.82	0.4880	1.16	0.8644	1.50	0.9778	1.84	0.9977	2.18	0.9999
0.49	0.0300	0.83	0.5038	1.17	0.8706	1.51	0.9791	1.85	0.9979	2.19	0.9999
0.50	0.0361	0.84	0.5194	1.18	0.8765	1.52	0.9803	1.86	0.9980	2.20	0.9999
0.51	0.0428	0.85	0.5347	1.19	0.8823	1.53	0.9815	1.87	0.9981	2.21	0.9999
0.52	0.0503	0.86	0.5497	1.20	0.8877	1.54	0.9826	1.88	0.9983	2.22	0.9999
0.53	0.0585	0.87	0.5645	1.21	0.8930	1.55	0.9836	1.89	0.9984	2.23	0.9999
0.54	0.0675	0.88	0.5791	1.22	0.8981	1.56	0.9846	1.90	0.9985	2.24	0.9999
0.55	0.0772	0.89	0.5933	1.23	0.9030	1.57	0.9855	1.91	0.9986	2.25	0.9999
0.56	0.0876	0.90	0.6073	1.24	0.9076	1.58	0.9864	1.92	0.9987	2.26	0.9999
0.57	0.0987	0.91	0.6209	1.25	0.9121	1.59	0.9873	1.93	0.9988	2.27	0.9999
0.58	0.1104	0.92	0.6343	1.26	0.9164	1.60	0.9880	1.94	0.9989	2.28	0.9999
0.59	0.1228	0.93	0.6473	1.27	0.9206	1.61	0.9888	1.95	0.9990	2.29	0.9999
0.60	0.1357	0.94	0.6601	1.28	0.9245	1.62	0.9895	1.96	0.9991	2.30	0.9999
0.61	0.1492	0.95	0.6725	1.29	0.9283	1.63	0.9902	1.97	0.9991	2.31	1.0000
0.62	0.1632	0.96	0.6846	1.30	0.9319	1.64	0.9908	1.98	0.9992		
0.63	0.1778	0.97	0.6963	1.31	0.9354	1.65	0.9914	1.99	0.9993		
0.64	0.1927	0.98	0.7079	1.32	0.9387	1.66	0.9919	2.00	0.9993		
0.65	0.2080	0.99	0.7191	1.33	0.9418	1.67	0.9924	2.01	0.9994		

附表 8 正交设计表

$L_4(2^3)$

试验号＼列号	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1
组	1	2	

$L_8(2^7)$

试验号＼列号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
组	1	2		3			

$L_8(2^7)$二列间的交互作用

列号＼列号	1	2	3	4	5	6	7
	(1)	3	2	5	4	7	6
		(2)	1	6	7	4	5
			(3)	7	6	5	4
				(4)	1	2	3
					(5)	3	2
						(6)	1
							(7)

$L_8(2^7)$表头设计

因子数＼列号	1	2	3	4	5	6	7
3	A	B	$A\times B$	C	$A\times C$	$B\times C$	
4	A	B	$A\times B$ $C\times D$	C	$A\times C$ $B\times D$	$B\times C$ $A\times D$	D
4	A	B $C\times D$	$A\times B$	C $B\times D$	$A\times C$	D $B\times C$	$A\times D$
5	A $D\times E$	B $C\times D$	$A\times B$ $C\times E$	C $B\times D$	$A\times C$ $B\times E$	D $A\times E$ $B\times C$	E $A\times D$

$L_8(4\times2^4)$

试验号 \ 列号	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	2	1	1	2	2
4	2	2	2	1	1
5	3	1	2	1	2
6	3	2	1	2	1
7	4	1	2	2	1
8	4	2	1	1	2

$L_8(4\times2^4)$表头设计

因子数 \ 列号	1	2	3	4	5
2	A	B	$(A\times B)_1$	$(A\times B)_2$	$(A\times B)_3$
3	A	B	C		
4	A	B	C	D	
5	A	B	C	D	E

$L_{12}(2^{11})$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1

$L_{16}(2^{15})$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1
组	1	2			3						4				

$L_{16}(2^{15})$二列间的交互作用

列号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	(1)	3	2	5	4	7	6	9	8	11	10	13	12	15	14
		(2)	1	6	7	4	5	10	11	8	9	14	15	12	13
			(3)	7	6	5	4	11	10	9	8	15	14	13	12
				(4)	1	2	3	12	13	14	15	8	9	10	11
					(5)	3	2	13	12	15	14	9	8	11	10
						(6)	1	14	15	12	13	10	11	8	9
							(7)	15	14	13	12	11	10	9	8
								(8)	1	2	3	4	5	6	7
									(9)	3	2	5	4	7	6
										(10)	1	6	7	4	5
											(11)	7	6	5	4
												(12)	1	2	3
													(13)	3	2
														(14)	1

$L_{16}(2^{15})$表头设计

因子数 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	A	B	$A\times B$	C	$A\times C$	$B\times C$		D	$A\times D$	$B\times D$		$C\times D$			
5	A	B	$A\times B$	C	$A\times C$	$B\times C$	$D\times E$	D	$A\times D$	$B\times D$	$C\times E$	$C\times D$	$B\times E$	$A\times E$	E
6	A	B	$A\times B$ $D\times E$	C	$A\times C$ $D\times F$	$B\times C$ $E\times F$		D	$A\times D$ $B\times E$ $C\times F$	$B\times D$ $A\times E$	E	$C\times D$ $E\times F$	F	$C\times E$ $B\times F$	
7	A	B	$A\times B$ $D\times E$ $F\times G$	C	$A\times C$ $D\times F$ $E\times G$	$B\times C$ $E\times F$ $D\times G$		D	$A\times D$ $B\times E$ $C\times F$	$B\times D$ $A\times E$ $C\times G$	E	$C\times D$ $A\times F$ $B\times G$	F	G	$C\times E$ $B\times F$ $A\times G$
8	A	B	$A\times B$ $D\times E$ $F\times G$ $C\times H$	C	$A\times C$ $D\times F$ $E\times G$ $B\times H$	$B\times C$ $E\times F$ $D\times G$ $A\times H$	H	D	$A\times D$ $B\times E$ $C\times F$ $G\times H$	$B\times D$ $A\times E$ $C\times G$ $F\times H$	E	$C\times D$ $A\times F$ $B\times G$ $E\times H$	F	G	$C\times E$ $B\times F$ $A\times G$ $D\times H$

$L_{16}(4\times2^{12})$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	2	2	2	2	2	2	2	2	1	1	1	1
5	2	1	1	2	2	1	1	2	2	1	1	2	2
6	2	1	1	2	2	2	2	1	1	2	2	1	1
7	2	2	2	1	1	1	1	2	2	2	2	1	1
8	2	2	2	1	1	2	2	1	1	1	1	2	2
9	3	1	2	1	2	1	2	1	2	1	2	1	2
10	3	1	2	1	2	2	1	2	1	2	1	2	1
11	3	2	1	2	1	1	2	1	2	2	1	2	1
12	3	2	1	2	1	2	1	2	1	1	2	1	2
13	4	1	2	2	1	1	2	2	1	1	2	2	1
14	4	1	2	2	1	2	1	1	2	2	1	1	2
15	4	2	1	1	2	1	2	2	1	2	1	1	2
16	4	2	1	1	2	2	1	1	2	1	2	2	1

$L_{16}(4\times2^{12})$表头设计

因子数 \ 列号	1	2	3	4	5	6	7
3	A	B	$(A\times B)_1$	$(A\times B)_2$	$(A\times B)_3$	C	$(A\times C)_1$
4	A	B	$(A\times B)_1$ $C\times D$	$(A\times B)_2$	$(A\times B)_3$	C	$(A\times C)_1$ $B\times D$
5	A	B	$(A\times B)_1$ $C\times D$	$(A\times B)_2$ $C\times E$	$(A\times B)_3$	C	$(A\times C)_1$ $B\times D$

因子数 \ 列号	8	9	10	11	12	13
3	$(A\times C)_2$	$(A\times C)_2$	$B\times C$			
4	$(A\times C)_2$	$(A\times C)_3$	$B\times C$ $(A\times D)_1$	D	$(A\times D)_3$	$(A\times D)_2$
5	$(A\times C)_2$ $B\times E$	$(A\times C)_3$	$B\times C$ $(A\times D)_1$ $(A\times E)_2$	D $(A\times E)_3$	E $(A\times D)_3$	$(A\times E)_1$ $(A\times D)_2$

$L_{16}(4^2 \times 2^9)$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	2	1	1	1	2	2	2	2	2	2
3	1	3	2	2	2	1	1	1	2	2	2
4	1	4	2	2	2	2	2	2	1	1	1
5	2	1	1	2	2	1	2	2	1	2	2
6	2	2	1	2	2	2	1	1	2	1	1
7	2	3	2	1	1	1	2	2	2	1	1
8	2	4	2	1	1	2	1	1	1	2	2
9	3	1	2	1	2	2	1	2	2	1	2
10	3	2	2	1	2	1	2	1	1	2	1
11	3	3	1	2	1	2	1	2	1	2	1
12	3	4	1	2	1	1	2	1	2	1	2
13	4	1	2	2	1	2	2	1	2	2	1
14	4	2	2	2	1	1	1	2	1	1	2
15	4	3	1	1	2	2	2	1	1	1	2
16	4	4	1	1	2	1	1	2	2	2	1

$L_{16}(4^3 \times 2^6)$

试验号 \ 列号	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	1	2	2	1	1	2	2	2	2
3	1	3	3	2	2	1	1	2	2
4	1	4	4	2	2	2	2	1	1
5	2	1	2	2	2	1	2	1	2
6	2	2	1	2	2	2	1	2	1
7	2	3	4	1	1	1	2	2	1
8	2	4	3	1	1	2	1	1	2
9	3	1	3	1	2	2	2	2	1
10	3	2	4	1	2	1	1	1	2
11	3	3	1	2	1	2	2	1	2
12	3	4	2	2	1	1	1	2	1
13	4	1	4	2	1	2	1	2	2
14	4	2	3	2	1	1	2	1	1
15	4	3	2	1	2	2	1	1	1
16	4	4	1	1	2	1	2	2	2

$L_{16}(4^4 \times 2^3)$

试验号 \ 列号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	2	2	2	1	2	2
3	1	3	3	3	2	1	2
4	1	4	4	4	2	2	1
5	2	1	2	3	2	2	1
6	2	2	1	4	2	1	2
7	2	3	4	1	1	2	2
8	2	4	3	2	1	1	1
9	3	1	3	3	1	2	2
10	3	2	4	4	1	1	1
11	3	3	1	2	2	2	1
12	3	4	2	1	2	1	2
13	4	1	4	2	2	1	2
14	4	2	3	1	2	2	1
15	4	3	2	4	1	1	1
16	4	4	1	3	1	2	2

$L_{16}(4^5)$

试验号 \ 列号	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	2	1	2	3	4
6	2	2	1	4	3
7	2	3	4	1	2
8	2	4	3	2	1
9	3	1	3	4	2
10	3	2	4	3	1
11	3	3	1	2	4
12	3	4	2	1	3
13	4	1	4	2	3
14	4	2	3	1	4
15	4	3	2	4	1
16	4	4	1	3	2
组	1	2			

$L_{16}(8\times2^8)$

试验号 \ 列号	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	1	2	2	2	2	2	2	2	2
3	2	1	1	1	1	2	2	2	2
4	2	2	2	2	2	1	1	1	1
5	3	1	1	1	1	1	1	2	2
6	3	2	2	2	2	2	2	1	1
7	4	1	1	1	1	2	2	1	1
8	4	2	2	2	2	1	1	2	2
9	5	1	2	1	2	1	2	1	2
10	5	2	1	2	1	2	1	2	1
11	6	1	2	1	2	2	1	2	1
12	6	2	1	2	1	1	2	1	2
13	7	1	2	2	1	1	2	2	1
14	7	2	1	1	2	2	1	1	2
15	8	1	2	2	1	2	1	1	2
16	8	2	1	1	2	1	2	2	1

$L_{20}(2^{19})$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1
3	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1	2
4	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1	2	2
5	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1	2	2	1
6	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1	2	2	1	1
7	2	2	2	1	2	1	2	1	1	1	1	2	2	1	2	2	1	1	2
8	2	2	1	2	1	2	1	1	1	1	2	2	1	2	2	1	1	2	2
9	2	1	2	1	2	1	1	1	1	2	2	1	2	2	1	1	2	2	2
10	1	2	1	2	1	1	1	1	2	2	1	2	2	1	1	2	2	2	2
11	2	1	2	1	2	1	1	2	2	1	2	2	1	1	2	2	2	2	1
12	1	2	1	1	1	1	2	2	1	2	2	1	1	2	2	2	2	1	2
13	2	1	1	1	1	2	2	1	2	2	1	1	2	2	2	2	1	2	1
14	1	1	1	1	2	2	1	2	2	1	1	2	2	2	2	1	2	1	2
15	1	1	1	2	2	1	2	2	1	1	2	2	2	2	1	2	1	2	1
16	1	1	2	2	1	2	2	1	1	2	2	2	2	1	2	1	2	1	1
17	1	2	2	1	2	2	1	1	2	2	2	2	1	2	1	2	1	1	1
18	2	2	1	2	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1
19	2	1	2	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2
20	1	2	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2

$L_9(3^4)$

试验号 \ 列号	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1
组	1		2	

注　任意二列间的交互作用为另外二列。

$L_{18}(2\times 3)^7$

试验号 \ 列号	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

$L_{27}(3^{13})$

试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3	3	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6	1	2	2	2	3	3	3	1	1	1	2	2	2
7	1	3	3	3	1	1	1	3	3	3	2	2	2
8	1	3	3	3	2	2	2	1	1	1	3	3	3
9	1	3	3	3	3	3	3	2	2	2	1	1	1
10	2	1	2	3	1	2	3	1	2	3	1	2	3
11	2	1	2	3	2	3	1	2	3	1	2	3	1
12	2	1	2	3	3	1	2	3	1	2	3	1	2
13	2	2	3	1	1	2	3	2	3	1	3	1	2
14	2	2	3	1	2	3	1	3	1	2	1	2	3
15	2	2	3	1	3	1	2	1	2	3	2	3	1
16	2	3	1	2	1	2	3	3	1	2	2	3	1
17	2	3	1	2	2	3	1	1	2	3	3	1	2
18	2	3	1	2	3	1	2	2	3	1	1	2	3
19	3	1	3	2	1	3	2	1	3	2	1	3	2
20	3	1	3	2	2	1	3	2	1	3	2	1	3
21	3	1	3	2	3	2	1	3	2	1	3	2	1
22	3	2	1	3	1	3	2	2	1	3	3	2	1
23	3	2	1	3	2	1	3	3	2	1	1	3	2
24	3	2	1	3	3	2	1	1	3	2	2	1	3
25	3	3	2	1	1	3	2	3	2	1	2	1	3
26	3	3	2	1	2	1	3	1	3	2	3	2	1
27	3	3	2	1	3	2	1	2	1	3	1	3	2
组	1	2			3								

$L_{27}(3^{13})$表头设计

因子数 \ 列号	1	2	3	4	5	6	7
3	A	B	$(A\times B)_1$	$(A\times C)_2$	C	$(A\times C)_1$	$(A\times C)_2$
4	A	B	$(A\times B)_1$ $(C\times D)_2$	$(A\times B)_2$	C	$(A\times C)_1$ $(B\times D)_2$	$(A\times C)_2$

因子数 \ 列号	8	9	10	11	12	13
3	$(B\times C)_1$			$(B\times C)_2$		
4	$(B\times C)_1$ $(A\times D)_2$	D	$(A\times D)_1$	$(B\times C)_2$	$(B\times D)_1$	$(C\times D)_1$

$L_{27}(3^{13})$二列间的交互作用

列号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13
	(1)	3	2	2	6	5	5	9	8	8	12	11	11
		4	4	3	7	7	6	10	10	9	13	13	12
		(2)	1	1	8	9	10	5	6	7	5	6	7
			4	3	11	12	13	11	12	13	8	9	10
			(3)	1	9	10	8	7	5	6	6	7	5
				2	13	11	12	12	13	11	10	8	9
				(4)	10	8	9	6	7	5	7	5	6
					12	13	11	13	11	12	9	10	8
					(5)	1	1	2	3	4	2	4	3
						7	6	11	13	12	8	10	9
						(6)	1	4	2	3	3	2	4
							5	13	12	11	10	9	8
							(7)	3	4	2	4	3	2
								12	11	13	9	8	10
								(8)	1	1	2	3	4
									10	9	5	7	6
									(9)	1	4	2	3
										8	7	6	5
										(10)	3	4	2
											6	5	7
											(11)	1	1
												13	12
												(12)	1
													11

$L_{25}(5^6)$

试验号 \ 列号	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	2	2	2	2	2
3	1	3	3	3	3	3
4	1	4	4	4	4	4
5	1	5	5	5	5	5
6	2	1	2	3	4	5
7	2	2	3	4	5	1
8	2	3	4	5	1	2
9	2	4	5	1	2	3
10	2	5	1	2	3	4
11	3	1	3	5	2	4
12	3	2	4	1	3	5
13	3	3	5	2	4	1
14	3	4	1	3	5	2
15	3	5	2	4	1	3
16	4	1	4	2	5	3
17	4	2	5	3	1	4
18	4	3	1	4	2	5
19	4	4	2	5	3	1
20	4	5	3	1	4	2
21	5	1	5	4	3	2
22	5	2	1	5	4	3
23	5	3	2	1	5	4
24	5	4	3	2	1	5
25	5	5	4	3	2	1
组	1	2				

附表 9　拉普拉斯变换简表*

	$f(t)$	$F(s)$
1	1	$\frac{1}{s}$
2	e^{at}	$\frac{1}{s-a}$
3	$t^m,(m>-1)$	$\frac{\Gamma(m+1)}{s^{m+1}}$
4	$t^m e^{at}(m>-1)$	$\frac{\Gamma(m+1)}{(s-a)^{m+1}}$
5	$\sin at$	$\frac{a}{s^2+a^2}$
6	$\cos at$	$\frac{s}{s^2+a^2}$
7	$\mathrm{sh}\,at$	$\frac{a}{s^2-a^2}$
8	$\mathrm{ch}\,at$	$\frac{s}{s^2-a^2}$
9	$t^m \sin at,(m>-1)$	$\frac{\Gamma(m+1)}{2j(s^2+a^2)^{m+1}}$ $\times[(s+ja)^{m+1}-(s-ja)^{m+1}]$
10	$t^m \cos at,(m>-1)$	$\frac{\Gamma(m+1)}{2(s^2+a^2)^{m+1}}$ $\times[(s+ja)^{m+1}+(s-ja)^{m+1}]$
11	$e^{-bt}\sin at$	$\frac{a}{(s+b)^2+a^2}$
12	$e^{-bt}\cos at$	$\frac{s+b}{(s+b)^2+a^2}$

* 取自《积分变换》(工程数学),高等教育出版社,1978

续表

	$f(t)$	$F(s)$
13	$e^{-bt}\sin(at+c)$	$\dfrac{(s+b)\sin c+a\cos c}{(s+b)^2+a^2}$
14	$\sin^2 t$	$\dfrac{1}{2}\left(\dfrac{1}{s}-\dfrac{s}{s^2+4}\right)$
15	$\cos^2 t$	$\dfrac{1}{2}\left(\dfrac{1}{s}+\dfrac{s}{s^2+4}\right)$
16	$\sin at\sin bt$	$\dfrac{2abs}{[s^2+(a+b)^2][s^2+(a-b)^2]}$
17	$e^{at}-e^{bt}$	$\dfrac{a-b}{(s-a)(s-b)}$
18	$ae^{at}-be^{bt}$	$\dfrac{(a-b)s}{(s-a)(s-b)}$
19	$\dfrac{1}{a}\sin at-\dfrac{1}{b}\sin bt$	$\dfrac{b^2-a^2}{(s^2+a^2)(s^2+b^2)}$
20	$\cos at-\cos bt$	$\dfrac{(b^2-a^2)s}{(s^2+a^2)(s^2+b^2)}$
21	$\dfrac{1}{a^2}(1-\cos at)$	$\dfrac{1}{s(s^2+a^2)}$
22	$\dfrac{1}{a^3}(at-\sin at)$	$\dfrac{1}{s^2(s^2+a^2)}$
23	$\dfrac{1}{a^4}(\cos at-1)+\dfrac{1}{2a^2}t^2$	$\dfrac{1}{s^3(s^2+a^2)}$
24	$\dfrac{1}{a^4}(\mathrm{ch}at-1)-\dfrac{1}{2a^2}t^2$	$\dfrac{1}{s^3(s^2-a^2)}$
25	$\dfrac{1}{2a^3}(\sin at-at\cos at)$	$\dfrac{1}{(s^2+a^2)^2}$
26	$\dfrac{t}{2a}\sin at$	$\dfrac{s}{(s^2+a^2)^2}$
27	$\dfrac{1}{2a}(\sin at+at\cos at)$	$\dfrac{s^2}{(s^2+a^2)^2}$
28	$\dfrac{1}{a^4}(1-\cos at)-\dfrac{1}{2a^3}t\sin at$	$\dfrac{1}{s(s^2+a^2)^2}$

	$f(t)$	$F(s)$
29	$(1-at)e^{-at}$	$\frac{s}{(s+a)^2}$
30	$t\left(1-\frac{a}{2}t\right)e^{-at}$	$\frac{s}{(s+a)^3}$
31	$\frac{1}{a}(1-e^{-at})$	$\frac{1}{s(s+a)}$
32①	$\frac{1}{ab}+\frac{1}{b-a}\left(\frac{e^{-bt}}{b}-\frac{e^{-at}}{a}\right)$	$\frac{1}{s(s+a)(s+b)}$
33①	$\frac{e^{-at}}{(b-a)(c-a)}+\frac{e^{-bt}}{(a-b)(c-b)}$ $+\frac{e^{-ct}}{(a-c)(b-c)}$	$\frac{1}{(s+a)(s+b)(s+c)}$
34①	$\frac{ae^{-at}}{(c-a)(a-b)}+\frac{be^{-bt}}{(a-b)(b-c)}$ $+\frac{ce^{-ct}}{(b-c)(c-a)}$	$\frac{s}{(s+a)(s+b)(s+c)}$
35①	$\frac{a^2e^{-at}}{(c-a)(b-a)}+\frac{b^2e^{-bt}}{(a-b)(c-b)}$ $+\frac{c^2e^{-ct}}{(b-c)(a-c)}$	$\frac{s^2}{(s+a)(s+b)(s+c)}$
36①	$\frac{e^{-at}-e^{-bt}[1-(a-b)t]}{(a-b)^2}$	$\frac{1}{(s+a)(s+b)^2}$
37①	$\frac{[a-b(a-b)t]e^{-bt}-ae^{-at}}{(a-b)^2}$	$\frac{s}{(s+a)(s+b)^2}$
38	$e^{-at}-e^{\frac{at}{2}}\left(\cos\frac{\sqrt{3}\,at}{2}-\sqrt{3}\sin\frac{\sqrt{3}\,at}{2}\right)$	$\frac{3a^2}{s^3+a^3}$
39	$\sin at\,\mathrm{ch}\,at-\cos at\,\mathrm{sh}\,at$	$\frac{4a^3}{s^4+4a^4}$
40	$\frac{1}{2a^2}\sin at\,\mathrm{sh}\,at$	$\frac{s}{s^4+4a^4}$
41	$\frac{1}{2a^3}(\mathrm{sh}\,at-\sin at)$	$\frac{1}{s^4-a^4}$

续表

	$f(t)$	$F(s)$
42	$\frac{1}{2a^2}(\text{ch}at-\cos at)$	$\frac{s}{s^4-a^4}$
43	$\frac{1}{\sqrt{\pi t}}$	$\frac{1}{\sqrt{s}}$
44	$2\sqrt{\frac{t}{\pi}}$	$\frac{1}{s\sqrt{s}}$
45	$\frac{1}{\sqrt{\pi t}}e^{at}(1+2at)$	$\frac{s}{(s-a)\sqrt{s-a}}$
46	$\frac{1}{2\sqrt{\pi t^3}}(e^{bt}-e^{at})$	$\sqrt{s-a}-\sqrt{s-b}$
47	$\delta(t)$	1
48[2]	$J_0(at)$	$\frac{1}{\sqrt{s^2+a^2}}$
49[2]	$I_0(at)$	$\frac{1}{\sqrt{s^2-a^2}}$
50	$J_0(2\sqrt{at})$	$\frac{1}{s}e^{-\frac{a}{s}}$
51	$\frac{1}{\sqrt{\pi t}}\cos 2\sqrt{at}$	$\frac{1}{\sqrt{s}}e^{-\frac{a}{s}}$
52	$\frac{1}{\sqrt{\pi t}}\text{ch}2\sqrt{at}$	$\frac{1}{\sqrt{s}}e^{\frac{a}{s}}$
53	$\frac{1}{\sqrt{\pi t}}\sin 2\sqrt{at}$	$\frac{1}{s\sqrt{s}}e^{-\frac{a}{s}}$
54	$\frac{1}{\sqrt{\pi t}}\text{sh}2\sqrt{at}$	$\frac{1}{s\sqrt{s}}e^{\frac{a}{s}}$
55	$\frac{1}{t}(e^{bt}-e^{at})$	$\ln\frac{s-a}{s-b}$
56	$\frac{2}{t}\text{sh}at$	$\ln\frac{s+a}{s-a}=2\text{Ar th}\frac{a}{s}$
57	$\frac{2}{t}(1-\cos at)$	$\ln\frac{s^2+a^2}{s^2}$

续表

	$f(t)$	$F(s)$
58	$\frac{2}{t}(1-\mathrm{ch}at)$	$\ln\frac{s^2-a^2}{s^2}$
59	$\frac{1}{t}\sin at$	$\mathrm{arctg}\ \frac{a}{s}$
60	$\frac{1}{t}(\mathrm{ch}at-\cos bt)$	$\ln\sqrt{\frac{s^2+b^2}{s^2-a^2}}$
61②	$\frac{1}{\pi t}\sin(2a\sqrt{t})$	$\mathrm{erf}\left(\frac{a}{\sqrt{s}}\right)$
62③	$\frac{1}{\sqrt{\pi t}}e^{-2a\sqrt{t}}$	$\frac{1}{\sqrt{s}}e^{\frac{a^2}{s}}\mathrm{erfc}\left(\frac{a}{\sqrt{s}}\right)$
63	$\mathrm{erfc}\left(\frac{a}{2\sqrt{t}}\right)$	$\frac{1}{s}e^{-a\sqrt{t}}$
64	$\mathrm{erf}\left(\frac{t}{2a}\right)$	$\frac{1}{s}e^{a^2s^2}\mathrm{erfc}(as)$

① 式中 a,b,c 为不相等的常数.

② $I_n(x)=j^{-n}J_n(jx)$，J_n 称为第一类 n 阶贝塞尔(Bessel)函数. I_n 称为第一类 n 阶变形的贝塞尔函数，或称为虚宗量的贝塞尔函数.

③ $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$，称为误差函数.

$\mathrm{erfc}(x)=1-\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-t^2}dt$，称为余误差函数.

参 考 文 献

1 复旦大学编．概率论，第一册．人民教育出版社，1979

2 王梓坤著．概率论基础及其应用．科学出版社，1976

3 W 费勒著，胡迪鹤等译．概率论及其应用（上册）．科学出版社，1964

4 W 费勒著，刘文译．概率论及其应用（下册）．科学出版社，1979

5 钟开来著．概率论教程．上海科技出版社，1989

6 严士健，王隽骧，刘秀芳著．概率论基础．1982

7 胡迪鹤著．分析概率论．科学出版社，1984

8 华东师范大学编．概率论与数理统计教程．高等教育出版社，1983

9 复旦大学编．概率论，第二册第一分册．人民教育出版社，1979

10 复旦大学编．概率论，第二册第二分册．人民教育出版社，1980

11 H 克拉美著，郑朴，吴锦译．统计学的数学方法．高等教育出版社，1960

12 数学研究所编．常用数理统计方法．科学出版社，1979

13 陈希孺著．数理统计引论．科学出版社，1981

14 陈希孺著．非参数统计．上海科学技术出版社，1989

15 上海师范大学数学系概率统计教研组编．回归分析及其试验设计．上海教育出版社，1978

16 陈希孺，王松桂编著．近代实用回归分析．广西人民出版社，

1983
17 王学仁,王松桂编译．实用多元统计分析．上海科学技术出版社,1990
18 张尧庭,方开泰著．多元统计分析引论．科学出版社,1982
19 M 肯德尔著,中国科学院数学研究所概率组译．多元分析．科学出版社，1983
20 中国科学院数学研究所统计组编．抽样检验方法．科学出版社,1978
21 许宝禄著．抽样论．北京大学出版社,1982
22 张尧庭,陈汉峰编著．贝叶斯统计推断．科学出版社,1991
23 中国科学院数学研究所概率统计室编．常用数理统计表．科学出版社,1979
24 王梓坤．随机过程论．科学出版社,1965
25 李漳南,吴荣．随机过程教程．高等教育出版社，1987
26 复旦大学．随机过程．(概率论第三册)
27 基赫曼,斯科罗霍德著,周概容,刘嘉焜译．随机过程论．科学出版社,1986
28 伊藤清著．刘璋temp译．随机过程．上海科学技术出版社,1961
29 巴鲁查-赖特著,杨纪珂,吴立德译．马尔可夫过程论初步及其应用．上海科学技术出版社,1979
30 陆大绘．随机过程及其应用．清华大学出版社
31 安鸿志等．时间序列的分析与应用．科学出版社，1983
32 闵华玲．随机过程．同济大学出版社,1987
33 雅格龙．平稳随机函数导论．数学进展,1956,2：1,3～153
34 卡罗柳科 B C,斯科罗霍特 A B 著．周概容译．随机过程论(上)．南开大学数学系,1980
35 杨向群．可列马尔科夫过程构造论．湖南科技出版社,1981

36　蒋庆琅，方积乾译．随机过程原理与生命科学模型．上海翻译出版公司，1987
37　哈尔莫斯，P R 著，王建华译．测度论．科学出版社
38　龚光鲁．随机微分方程引论．北京大学出版社，1987
39　帕尔逊，E 著，邓永录，杨振明译．随机过程．高等教育出版社，1987
40　严加安．鞅与随机积分引论．上海科学技术出版社，1981
41　弗里德曼著，吴让泉译．随机微分方程及其应用．科学出版社，第一卷 1983，第二卷 1986
42　司曲斯 F 著，刘永才等译．随机微分方程理论及其应用．上海科学技术文献出版社，1986
43　安鸿志，陈兆国，杜金观，潘一民著．时间序列分析及其应用．科学出版社，1983
44　韩崇昭，王月娟，万百五．随机系统理论．西安交通大学，1981
45　徐光辉．随机服务系统．科学出版社，1980
46　王梓坤．生灭过程与马尔可夫链．科学出版社，1980
47　曹晋华，程侃．可靠性数学引论．科学出版社，1986
48　茆诗松，王玲玲．可靠性统计．华东师范大学出版社，1984
49　戴树森等．可靠性试验及其统计分析．国防工业出版社，1983
50　郦萌．计算机软件的可靠性．国防工业出版社，1988
51　布斯连科，НП，施廖盖尔，Ю А 著，徐毓云，杜淑敏译．统计试验法(蒙特卡洛法)．科学出版社，1964
52　彭克斯，JJ．卡森著，侯炳辉，张金水译．离散事件系统模拟，清华大学出版社，1988
53　董泽清．马尔可夫决策规划．中国科学院应用数学研究所讲义，1981

54 张万祺,姜启源,林元烈．考虑保证率约束的马氏决策规划和罚因子法．清华大学学报,1983,23(1),61—72

55 董泽清．刘克．无界报酬折扣半马尔可夫决策规划最优策略的结构．中国科学(A 辑)1985,11,975—985

56 林元烈．连续时间首达目标模型(I)—折扣矩最优模型．应用数学学报,1991,14(1),115—124

57 霍华德,R A 著,李为政等译．动态规划与马尔可夫过程．上海科学技术出版社,1960

58 Shiryayev A N. Probability Theory. Springer-Verlag,1984

59 Chow Y S. H Teicher. Probability Theory. Springer-Verlag,1978

60 Berger J O. Statistical Decision Theory. Springer-Verlag, 1980

61 Bajpal A C, Calus I M,Fairley J A. Statistical Methods for Engineers and Scientists. John Wiley & Sons, 1978

62 Kendall M,Staurt A. The Advanced Theory of Statistics. Vol. 2, Charles,1979

63 Lehman E L. Testing Statistical Hypothesis, 2nd ed. John Wiley & Sons,1986

64 Daniel W W. Applied Nonparametric Statistics. PWS-KENT,1990

65 Weisberg S. Applied Linear Regression. John Wiley & Sons,1985

66 Giri N. Analysis of Variance. South Asian Publishers,1986

67 Morrison D F. Applied Linear Statistical Methods. Prentice-Hall,1983

68 Johnson R A, Wichern D W. Applied Multivariate Statistical Analysis. Prentice-Hall,1982

69 Karlin S, Taylor H M. A First Course in Stochastic Processes. New York, Academic Press, 1975

70 Karlin S, Taylor H M. A Second Course in Stochastic Processes. New York, Academic Press, 1981

71 Ross S M. Stochastic Processes. John Wiley & Sons, 1983

72 Box G E P & Jenkins G M. Time Series Analysis forecasting and control. 1969

73 Priestley M B. Spectral Analysis and Time Series. Academic Press, 1981

74 Iosifescu M, Tǎutu P. Stochastic Processes and Applications in Biology and Medicine I theory. Springer-Verlag, 1973

75 Ethier S N & Kurtz T G. Markov Processes, Characterization and Convergence. John Wiley & Sons, 1986

76 Дынкин Е Б. Марковские Процессы. Москва, 1963

77 Litwiniszyn T. Application of the Equation of Stochastic Processes to Mechanics of Loose Bodies. Arch. Mech. Stos., Vol. 8, 393-411, 1956

78 Strook D W & Varadhan S R S. Multidimensional Diffusion Processes. New York, Spring-Verlag, 1979

79 Blumenthal R M & Getoor R K. Markov Processes and Potential theory. New York, Acad. Press, 1968

80 Chung K L. Lectures from Markov Processes to Brownian motion. New York, Springer-Verlag, 1982

81 Hunt G A. Markoff Processes and Potentials Ⅰ, Ⅱ, Ⅲ. Illinois J. Math. 1, 44-93, 316-69(1957) 2, 151-213(1958)

82 Port S C, & Stone C J. Brown Motion and Classical

Potential theory. Acad Press,1978

83 Athreya K B, Ney P E. Branching Process. Springer-Verlag,1970

84 Asmusen S,Hering H. Branching Process. Birknäuser,1983

85 Dellacherie C et Meyer P A. Probabilités et potentials. Vol. 2 Théorie des matingales. Paris, Hermann,1980

86 Doob J L. Semimartingales and subharmonic functions. Trams. Amer. Math. Soc. 77,86-121,1954

87 Koch G & Spreij P J C. Software reliability as an application of martingale & filtering theory. IEEE Trans. on reliability, Vol. R-32,No. 4,1983

88 Kumar P R & Varaiya P. Stochastic Systems: estimation, identification,and adaptive Control. Prentice-Hall,1986

89 Королюк В С, Скороход А В. Теория стохастических процессов. Справочник Вероятностный Теорий и Математической статистики. Москва,1978

90 Дынкин Е Б. Марковские Процессы. Москва,1963

91 Leondes C T. Control and Dynamie Systems. Acad,press, 1986

92 Wong E. Stochastic Processes in Infomation and Dynamical Systems. McGraw-Hill,1971

93 Kleinrock L. Queuing Systems,Vol I: Theory. John Wiley & Sons,1975

94 Cohen J W. The Single Server Queue. North-Holland,1982

95 Chaudhry M L & Templeton J G C. A first Course in Bulk Queues. John wiley & Sons,1983

96 Borovkov A A. Asymptotic Methods in Queuing Theory. translated by D. Newton, John Wiley & Sons,1984

97 Barlow R E & Proschan F . Mathematical Theory of Reliability. John Wiley & Sons,1965

98 Barlow R E. Statistical Theory of Reliability and Life Testing, To BEGIN WITH. Silver Spring, MD,1981

99 Lawless J E. Statistical Models and Method for Life Time Data. John Wiley & Sons,1982

100 Musa J D, A Iannino and Okumoto K. Software Reliability. McGraw-Hill,1987

101 Барзилович Е Ю и Д р. Вопросы Математической Теории Надёжности. Радио и связь,1983

102 Гнеденко Б В, Беляев Ю К и Соловьев А Д . Математические Методы в Теории Надёжности. Москва,Наука,1965

103 Baukuneeht K and Nef W. Digital Simulation. Springer-Verlag,1971

104 Fishman G S. Principles of Discrete event Simulation. New York, Wiley,1978

105 Graybeal W J and Pooch U W. Simulation: Principles and Methods. Winthrop,Cambriage, Mass,1980

106 Jansson B. Random Number Generators. Stocholm,1966

107 Kohlao J. Monte Carlo Simulation in Operation Research. Springer-Verlag,1972

108 Law A M and Kelton W D. Simulation Modeling and Analysis. New York, McGraw-Hill,1982

109 Бусленко Н П и Голенко Д И, Соболь И М, Сранович В Г. Метод Статистических Испытаний (Метод Монте Карло). Ризматгиз,1962

110 Blackwell D. Discrete Dynamic Programming. Ann. Math. Statist. Vol. 33,719-726,1962

111 Derman C. Finite State Markov Decision Processes. Acad. Press, 1970

112 Dynkin E B. Controlled Markov Process. Spring-Verlag, 1979

113 Ross S M. Introduction to Stochastic Dynamic Programming. Acad. Press, 1983

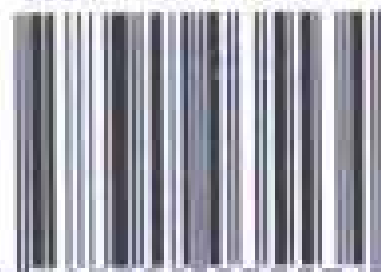
114 Monahan G A Survey of Partially Observable Markov Decision Processes: Theory, Models and Algorithms. Management Science. Vol. 28, No. 1, 1-13, 1982

数学科学的成就已成为当今高科技时代进步发展的重要基础，应用数学的发展是科技工业兴旺发达的强有力支柱。为了迎接21世纪的挑战，本手册向您介绍现代应用数学的各个分支，它们在解决科研、教学以及生产实践的各种问题中，提供不可缺少的工具。全书共计六卷：

《运筹学与最优化理论卷》
《现代应用分析卷》
《概率统计与随机过程卷》
《离散数学卷》
《分析与方程卷》
《计算与数值分析卷》

各卷内容自成体系，互相独立，方便读者按需选用。

ISBN 7-302-03553-9



定价：48.00 元